THE APPLICATION OF RESULTS IN FLUCTUATION AND RENEWAL THEORY TO QUEUEING PROBLEMS

A thesis for the degree

of

Master of Science in Statistics

Cheong Choong Kong

Department of Statistics
School of General Studies
Australian National University

April, 1965

CONTENTS

CONTENTS

CHAPTER I

Introduction and Summary.

The chief object of this thesis is to examine the results of renewal and fluctuation theory and apply them to some queueing problems, particularly those concerning waiting times.

Chapter II contains results that are a generalization of Cox's (1962) for the alternating renewal process. Cox considered a process $x(t)$ which is at any instant of time in exactly one of two alternating states, the lengths of stay in each state being independent random variables. The cyclic-k-state renewal process treated here is one which may be in one of k $(k \geq 2)$ distinct states, the occurrences of which follow one another in cyclic order. Unlike Cox's process, the lengths of stay in distinct states are not necessarily mutually independent. This process differs from W.L. Smith's (1955) semi-Markov processes in that not only are the 'waits' in distinct states not independent random variables, but a knowledge of the present state of the process is always sufficient to determine the subsequent state. Nevertheless, as will be shown in Chapter II, the cyclic-k-state process behaves in exactly the same manner as Cox's alternating process or Smith's semi-Markov process for large values of t. Taking the busy and idle periods of a single-server queue as the states of the process ($k = 2$), our results may be used to obtain limiting probabilities of the virtual waiting time being zero. The use of Cox's and Smith's results is limited to those queues in which the lengths of the busy and idle periods form independent random variables, e.g. queues with a Poisson input. By removing the assumption of independence between the lengths of time spent in distinct states, we make our results applicable to the general single-server queue.

Chapter III is a purely expository chapter and contains no original work. An attempt is made to review the known results of fluctuation theory. The two main methods of approach, the combinatorial and the analytic, are discussed, and proofs of major theorems sketched to illustrate them. § 3 of Chapter III is almost entirely confined to a discussion of the greater part of Chapter IV of Spitzer's (1964) book, in which a study of random walk on the real line by simple analytic methods leads to the basic results of fluctuation theory.

Chapter III serves in a preparatory capacity for Chapter IV. In the latter chapter, fluctuation theory results are used to derive exponential-type identities for the waiting time, idle time and busy period distributions of a general single-server queue; and further, to deduce the now familiar result of Lindley (1952) that with probability one, the waiting time of the nth customer, w_n, converges in distribution to a finite variable if and only if the mean inter-arrival time is strictly greater than the mean service time. An important result of this chapter relies, however, not on fluctuation theory, but on well-known renewal theory arguments. This result relates the equilibrium waiting time in an ergodic queue to moments of the busy period distribution (when these exist) and depends for its proof on only the renewal theorem of Feller (1949) and the recurrent property of the epochs of commencement of a busy period.

Chapter V is restricted to the ergodic queue. Here, the rate of convergence of the waiting time distribution to its limit is studied, by methods similar to those employed by Heathcote (1964). With the aid of the exponential identities derived in Chapter IV, it is shown that the waiting time distribution converges to its limit exponentially fast. In the course of proving this, other results are also obtained which are not devoid of independent interest.

The Laplace - Stieltjes transform of a function $B(x)$, $\int_{-\infty}^{\infty} e^{-sx}\, dB(x)$, will be denoted throughout the thesis by $B^*(s)$. Other notation and terms will be defined as soon as they are introduced.

CHAPTER II

The Cyclic-k-state Process

§ 1. Preliminary Remarks.

In this chapter, we consider a stochastic process $z(t)$ which is at any moment in exactly one of a finite number of states, to be denoted by the integers 1, 2, ..., k; and in which transitions from one state into another occur in a cyclic order ($1 \to 2, \ldots, i \to i+1, \ldots, k \to 1$). In this process, the lengths of stay in distinct states are not necessarily mutually independent, so that we are not in the class of semi-Markov processes. To simplify matters, we employ the very useful device of imbedding in $z(t)$ a sequence of regeneration points. A natural choice for such a sequence are the instants of transition from state k into state 1.

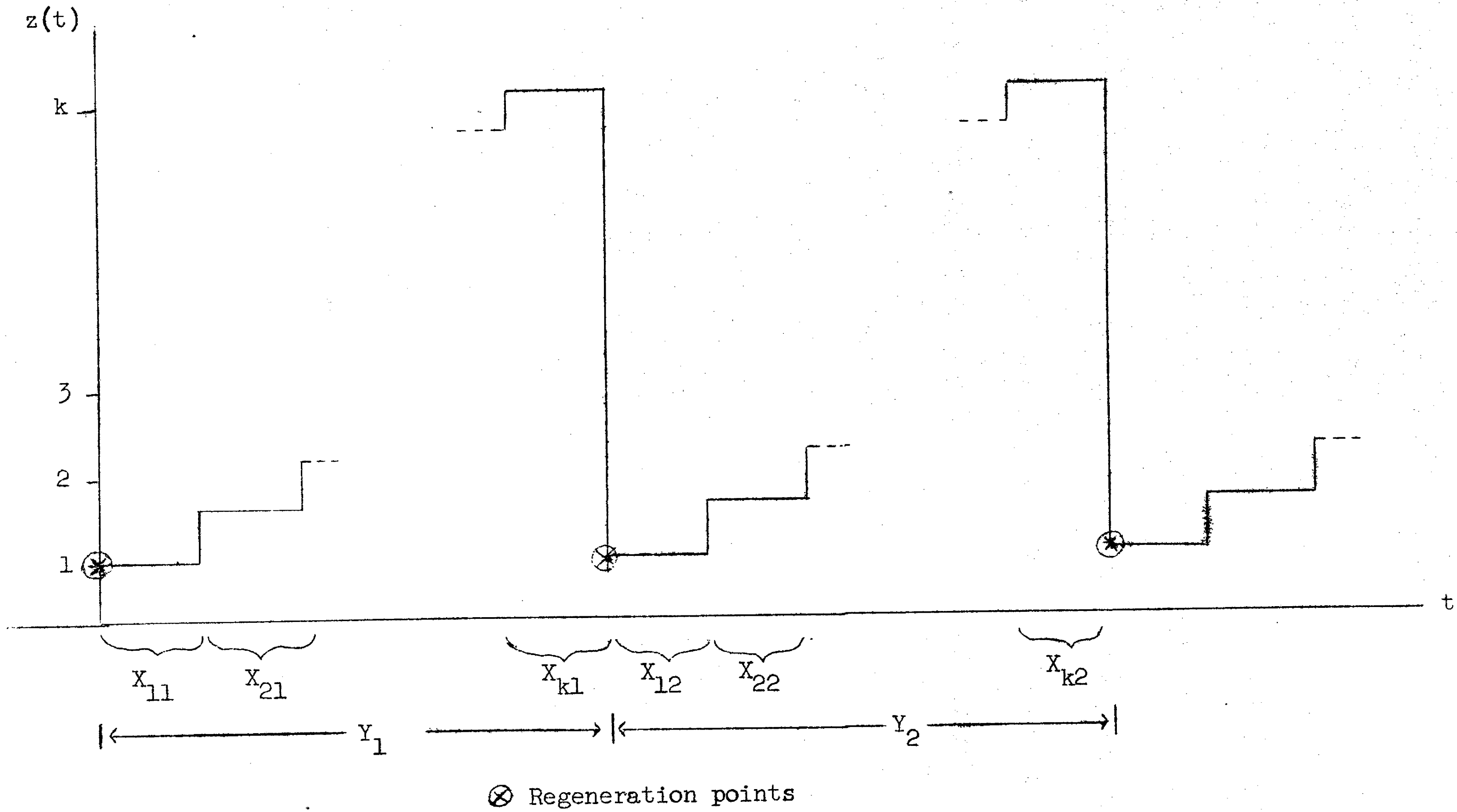
Such a process we call, for want of a better name, a "cyclic-k-state process".

The main results obtained here are analogous to those of Smith's [1955] for semi-Markov processes. Comparison between the two classes of results will be made in § 2.

Results for the alternating renewal process ($k = 2$ and $z(t)$ semi Markov) were obtained by Cox [1962]. These will emerge as special cases of the results derived in this chapter.

Suppose $z(t)$ is in state i for the μth time since $t = 0$ and denote the length of time spent in this state before the next transition by $X_{i\mu}$ ($i = 1, 2, \ldots, k$). We assume that each of $X_{1\mu}$, $X_{2\mu}$, ..., $X_{k\mu}$ is independent of $X_{i\nu}$ ($i = 1, \ldots, k$; $\nu < \mu$) i.e. the process "regenerates" after every transition into state 1. We also assume that for each μ, the random variable $X_{i\mu}$ has the same distribution as the random variable X_i, and $P[X_i \leq x] = X_i(x)$. Now let $Y_\mu = \sum_{i=1}^{k} X_{i\mu}$. Then $\{Y_\mu\}_{\mu=1}^{\infty}$ is a sequence of independent, random variables with a common distribution which we shall denote by $P[Y \leq x] = F(x)$. Let the number of transitions from state i into state $i+1$ (modulo k) in the time interval $[0, t]$ be a random variable $N_i(t)$, with expectation $E\, N_i(t) = H_i(t)$, and let $P[z(t) = i] = \pi_i(t)$ be the probability of the event that the process is in state i. Finally, for the sake of simplicity, we put $z(0) = 1$.

FIGURE A.



In § 2, we prove that $\lim_{t\to\infty} \pi_i(t)$ exists; in fact, this limit is $\frac{E[X_i]}{E[X_1 + \dots + X_k]}$ (and is interpreted as zero if $E[X_1 + \dots + X_k]$ is infinite). This result is then compared with a parallel one of Smith's on semi-Markov processes. Also, we obtain asymptotic expressions for $H_i(t)$.

In § 3, we consider the use of our results in the investigation of the probabilities of being in the busy and idle periods in queueing theory. It will be seen that the removal of the assumption of independence between the lengths of distinct states enables us to apply our results to the general single-server queue.

* * * * * * *

§ 2. Asymptotic Behaviour of $z(t)$

We begin by making the following definitions.

$$U_{o\mu} = 0, \qquad U_{i\mu} = X_{1\mu} + X_{2\mu} + \dots + X_{i\mu} \qquad i = 1, \dots, k; \quad \mu = 1, 2, \dots$$

$$S_o = 0, \qquad S_j = Y_1 + Y_2 + \dots + Y_j \qquad j = 1, 2, \dots$$

Note that, by the independence of the Y_i's, the distribution of S_j is the j-fold convolution of $F(x)$ with itself, and $U_{i\mu}$ has the same distribution which we denote by $\psi_i(x)$, for all μ.

Since the events $\{N_i(t) \geq j\}$ and $\{S_{j-1} + U_{ij} \leq t\}$ are equivalent,

$$H_i(t) = \sum_{j=1}^{\infty} j\, P[N_i(t) = j] = \sum_{j=1}^{\infty} j \left\{ P[N_i(t) \geq j] - P[N_i(t) \geq j+1] \right\}$$

$$= \sum_{j=1}^{\infty} P[S_{j-1} + U_{ij} \leq t] \quad .$$

Taking Laplace-Stieltjes transforms, and using the fact that S_{j-1} and U_{ij} are independent, we get

$$H_i^*(s) = \sum_{j=1}^{\infty} F_{j-1}^*(s)\, \psi_i^*(s)$$

$$= \frac{\psi_i^*(s)}{1 - F^*(s)} \quad , \qquad i = 1, 2, \dots, k \qquad (2.1)$$

This is the first of many examples of the usefulness of the imbedded regeneration points in $z(t)$.

By inverting (2.1), $H_i(t) = \psi_i(t) + \int_0^t F(t-u)\, dH_i(u)$.

$$(2.2)$$

Now,

$$\pi_i(t) = \sum_{j=1}^{\infty} P\left[S_{j-1} + U_{i-1,j} \leq t < S_{j-1} + U_{i,j}\right]$$

$$= \sum_{1}^{\infty} \int_0^t P\left[x + U_{i-1,j} \leq t < x + U_{i,j}\right] dP\left[S_{j-1} \leq x\right].$$

Let

$$G_{ij}(t) = P\left[U_{i-1,j} \leq t < U_{i,j}\right] .$$

Then

$$G_{ij}(t) = P\left[U_{i-1,j} \leq t\right] - P\left[U_{i,j} \leq t\right] , \qquad \text{where}$$

$$P\left[U_{o,1} \leq t\right] = \psi_o(t) = 1, \quad t \geq 0.$$

Hence

$$\pi_i^*(s) = \sum_{1}^{\infty} G_{ij}^*(s) \quad \{F^*(s)\}^{j-1}$$

$$= \frac{\psi_{i-1}^*(s) - \psi_i^*(s)}{1 - F^*(s)} \qquad i = 1, \ldots, k \qquad (2.3)$$

From (2.1) and (2.3),

$$\pi_i^*(s) = H_{i-1}^*(s) - H_i^*(s) \qquad i = 1, \ldots, k \qquad (2.4)$$

where $H_o^*(s) = \dfrac{\psi_o^*(s)}{1-F^*(s)} = \dfrac{1}{1-F^*(s)}$.

By inverting (2.3)

$$\pi_i(t) = \psi_{i-1}(t) - \psi_i(t) + \cancel{\int_0^t F(t-u)\, d\pi_i(u)}, \int_0^t \pi_i(t-u)\, dF(u)$$

$$i = 1, \ldots, k \quad (2.5)$$

Before proceeding to the first theorem, we define

$$a_i = E[X_i], \quad \mu_1 = E[Y] = \sum_{i=1}^{k} a_i \quad \text{and} \quad \mu_i = E[Y]^i .$$

<u>THEOREM 1</u>. For each $i = 1, 2, \ldots, k$, if $F(t)$ is not periodic and $F(\infty) = 1$, then as $t \to \infty$:

(i) $\dfrac{H_i(t)}{t} \to \dfrac{1}{\mu_1}$;

(ii) if $a_j < \infty$ for all $j \le i$, then $\pi_i(t) \to \dfrac{a_i}{\mu_1}$; and

(iii) if $\mu_2 < \infty$, then $H_i(t) - \dfrac{t}{\mu_1} \to \dfrac{\mu_2}{2\mu_1^2} - \dfrac{a_1 + a_2 + \cdots + a_i}{\mu_1}$;

where $\dfrac{1}{\mu_1}$ is interpreted as zero if $\mu_1 = \infty$.

Note: (a) The limits depend only on the expectations of the distributions; the "form" of the distributions play no part.

(b) The probability of being in state i at infinity is simply the ratio of the mean length of stay in state i to the mean length of the interval between two successive regeneration points; this is so even if the X_i's are not mutually independent.

Proof: (i) and (iii) follow directly from (2.1) and applications of the renewal theorems of (Smith, 1954).

(ii), the most interesting result of the three, requires for its proof an application of the key-renewal theorem (Smith, 1954).

Let

$$K^*(s) = \frac{F^*(s)}{1-F^*(s)} \quad .$$

Then from (2.3)

$$\pi_i^*(s) = \left\{\psi_{i-1}{}^*(s) - \psi_i^*(s)\right\}\left\{1 + K^*(s)\right\} \quad , \qquad \text{and}$$

$$\pi_i(t) = \psi_{i-1}(t) - \psi_i(t) + \int_0^t \left\{\psi_{i-1}(t-u) - \psi_i(t-u)\right\} dK(u).$$

Since $\psi_{i-1}(u) - \psi_i(u)$ is clearly of bounded variation, Lebesgue integrable on $(0, \infty)$ ($a_j < \infty$ for all $j \leq i$) and zero for negative arguments,

$$\int_0^t \left\{ \psi_{i-1}(t-u) - \psi_i(t-u) \right\} dK(u) \to \frac{1}{\mu_1} \int_0^\infty \left\{ \psi_{i-1}(u) - \psi_i(u) \right\} du$$

by the key-renewal theorem.

Hence,

$$\pi_i(t) \to \frac{(a_1 + a_2 + \cdots + a_i) - (a_1 + a_2 + \cdots + a_{i-1})}{\mu_1}$$

$$= \frac{a_i}{\mu_1} \qquad (= 0 \text{ if } \mu_1 = \infty).$$

It will be of interest to compare Theorem 1 with a parallel result of W.L. Smith. But first, a brief review of part of Smith's paper [1955] is necessary.

Let $\{t_i\}$, $i = 1, 2, \ldots$ be a renewal process with the common distribution function $G(t)$ and mean ν_1, $T_k = \sum_1^k t_i$ and N_t the greatest integer k such that $T_{k-1} \leq t$. Let $x(t)$ be a stochastic process taking values on Ω, an abstract space of elements $\underset{\sim}{x}$, and let A be some $\underset{\sim}{x}$-set. Smith defines a regenerative process $x(t)$ as one for which there exists a function ϕ_A, depending only on A, such that

$$P\left[x(t) \in A / x(0);\ N_t > 0;\ T_{N_t};\ x(t'),\ t' \leq T_{N_t} \right] = \phi_A(t - T_{N_t})$$

is a valid representation of the conditional probability for some $\{t_i\}$. Roughly, $x(t)$ is regenerative if for some renewal process $\{t_i\}$, $x(t)$ is independent of that part of its history prior to the last T_n. Next, Smith proceeds to prove (Theorem 2 of his paper) that if $x(t)$ is regenerative, then, under certain conditions,

$$\lim_{t\to\infty} P\left[x(t) \in A/x(0)\right] = \frac{1}{\nu_1}\int_0^\infty \phi_A(u)\left\{1-G(u)\right\} du \text{ if } \nu_1<\infty$$

$$= 0 \text{ if } \nu_1 = \infty.$$

Suppose $x(t)$ at any instant is in exactly one of a countable number of "states" (Ω is here the set of positive integers, and the sets A, values of $x(t)$, each consists of exactly one element.) Smith calls this process a semi-Markov process if:

(i) the sequence of values ("states") of $x(t)$, $\{x_n\}_1^\infty$ forms a Markov chain with transition probabilities $P\left[x_{n+1} = j/x_n = i\right] = p_{ij}$;

(ii) the lengths of stay in distinct states are independent random variables; and

(iii) none of these random variables is 0 with probability 1.

Now, for a given i, let R_i be the event "a transition into state i" and $\{t_j\}_1^\infty$ the intervals separating successive occurrences of R_i. Then clearly $\{t_j\}$ is a sequence of regeneration points and $x(t)$ is a regenerative process. Let X_i be the length of stay in state i, $a_i = EX_i$ and $\nu_1 = E[t_j]$.

Then, with the help of his Theorem 2, Smith proves (Theorem 5) that if R_i is not periodic, $a_i < \infty$ and a transition occurred at $t = 0$, then

$$\lim_{t\to\infty} P\left[x(t) = i/x(o)\right] = \frac{Q_i a_i}{\nu_1} \qquad , \text{ where}$$

Q_i is the probability that, given $x(0)$, R_i occurs in a finite time. If $\nu_1 = \infty$, then the limit is zero.

For the cyclic-k-state process $z(t)$, Q_i is clearly one if $F(\infty) = 1$, and so we have in part (ii) of Theorem 1 an analogue of Smith's Theorem 5.

<u>Example 2.1</u>

$k = 2$:

From (2.1) and (2.3),

$$H_1^*(s) = \frac{X_1^*(s)}{1-F^*(s)} \quad , \qquad H_2^*(s) = \frac{F^*(s)}{1-F^*(s)} \quad ,$$

$$\pi_1^*(s) = \frac{1-X_1^*(s)}{1-F^*(s)} \quad , \qquad \pi_2^*(s) \quad \frac{X_1^*(s) - F^*(s)}{1-F^*(s)} \quad ,$$

and it is evident that $\pi_1(t) + \pi_2(t) = 1$ as expected, and $\pi_1(t) = 1 + H_2(t) - H_1(t)$. Cox's results [1962] for the alternating renewal process may be obtained by assuming independence between X_1 and X_2 and replacing $F^*(s)$ in each of the equations above with $X_1^*(s)\, X_2^*(s)$. Note that the assumption of independence is unnecessary for deriving the interesting result

$\pi_1(t) = 1 + H_2(t) - H_1(t)$.

Suppose now that X_1 and X_2 are independent and negative exponentially distributed.

Let
$$X_1(x) = 1 - e^{-\lambda_1 x} \quad \text{and} \quad X_2(x) = 1 - e^{-\lambda_2 x} .$$

Then
$$a_i = \frac{1}{\lambda_i} , \quad \mu_1 = \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} , \quad \mu_2 = E\left[X_1 + X_2\right]^2 = 2\left(\frac{\lambda_1^2 + \lambda_2^2 + \lambda_1\lambda_2}{\lambda_1^2 \lambda_2^2}\right)$$

$$H_1^*(s) = \frac{\dfrac{\lambda_1}{\lambda_1 + s}}{1 - \dfrac{\lambda_1}{\lambda_1 + s} \cdot \dfrac{\lambda_2}{\lambda_2 + s}} = \left\{\frac{\lambda_1^2}{\lambda_1 + \lambda_2}\right\}\left\{\frac{1}{s+(\lambda_1 + \lambda_2)}\right\} + \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)s}$$

Inverting,
$$H_1(t) = \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} t - \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 e^{-(\lambda_1 + \lambda_2)t} + \text{constant}$$

As
$$H_1(0) = 0, \quad \text{constant} = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 .$$

Hence
$$H_1(t) = \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} t + \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 \left(1 - e^{-(\lambda_1 + \lambda_2)t}\right) ,$$

and
$$\frac{H_1(t)}{t} \to \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} = \frac{1}{\mu_1} .$$

Also,

$$H_1(t) - \frac{t}{\mu_1} = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 \left(1 - e^{-(\lambda_1 + \lambda_2)t}\right)$$

$$\to \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) = \frac{\mu_2}{2\mu_1^2} - \frac{a_1}{\mu_1}$$

$$\pi_1^*(s) = \frac{1 - \dfrac{\lambda_1}{\lambda_1 + s}}{1 - \dfrac{\lambda_1}{\lambda_1 + s} \cdot \dfrac{\lambda_2}{\lambda_2 + s}} = 1 - \frac{\lambda_1}{s + (\lambda_1 + \lambda_2)} .$$

Inverting,

$$\pi_1(t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2)t} + \text{constant} .$$

As

$$\pi_1(0) = 1, \quad \text{constant} = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

Hence
$$\pi_1(t) = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2)t}$$

$$\to \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{a_1}{\mu_1} .$$

Theorem 1 is restricted to the case ~~that~~ where $F(x)$ is an honest distribution function. If one of the X_i's, X_m, say, has a positive probability of being infinite, then for each $i = 1, \ldots, k$, $H_i(t)$ approaches a finite limit, the value of which depends on whether

i is strictly less than m or not, and the expected result,

$\lim_{t\to\infty} \pi_i(t) = \delta_{im}$, follows. All this is contained in the more general

<u>THEOREM 2</u>. Suppose $A = \left\{ j/P\,[X_j < \infty] < 1 \right\}$ is a non-empty sub-set of $\{1, 2, \ldots, k\}$ and $m = \min\,\{i/i \in A\}$. Then

$$\lim_{t\to\infty} H_i(t) = \begin{cases} \dfrac{\psi_i(\infty)}{1-F(\infty)} & i \geq m \\ \\ \dfrac{1}{1-F(\infty)} & i < m, \end{cases} \qquad \text{and}$$

$$\lim_{t\to\infty} \pi_i(t) = \begin{cases} 0 & i \notin A \\ \\ \dfrac{\psi_{i-1}(\infty) - \psi_i(\infty)}{1-F(\infty)} & i \in A \end{cases}$$

where $\psi_o(\infty) = 1$.

Proof: Since $H_i(t)$ is non-decreasing, a Tauberian argument may be applied to (2.1) to give

$$\lim_{t\to\infty} H_i(t) = \frac{\lim\limits_{s\to o} \psi_i^*(s)}{\lim\limits_{s\to o} [1-F^*(s)]} = \frac{\psi_i(\infty)}{1-F(\infty)} \quad .$$

To prove the second part of the theorem, recall (2.4)

$$\pi_1^*(s) = \frac{1 - \psi_1^*(s)}{1-F^*(s)} \quad , \quad \pi_i(t) = H_{i-1}(t) - H_i(t) \quad i = 2, 3, \ldots, k.$$

We divide the proof into the four exhaustive cases

(a) $1 < i < m$, (b) $i > m$, (c) $i = m > 1$ (d) $i = 1 \leq m$.

(a) $1 < i < m$: $\pi_i(t) \to 0$ from (2.4) and the first part of the theorem.

(b) $i > m$: $\pi_i(t) \to \dfrac{\psi_{i-1}(\infty) - \psi_i(\infty)}{1 - F(\infty)}$.

If $i \notin A$, then $\psi_{i-1}(\infty) = \psi_i(\infty)$ and $\pi_i(t) \to 0$.

(c) $i = m > 1$: $\pi_m(t) = H_{m-1}(t) - H_m(t)$

$$\to \frac{1}{1-F(\infty)} - \frac{\psi_m(\infty)}{1-F(\infty)} .$$

(d) $i = 1 \leq m$: (2.4) shows that $\sum_{i=1}^{k} \pi_i(t) = 1$. Since

$\lim_{t\to\infty} \pi_i(t)$ exists for all $2 \leq i \leq k$, the limit

of $\pi_1(t)$ exists as $t \to \infty$. Hence, by an Abelian argument

$$\lim_{t\to\infty} \pi_1(t) = \lim_{s\to o} \frac{1-\psi_1^*(s)}{1-F^*(s)} = \frac{1-\psi_1(\infty)}{1-F(\infty)} .$$

The proof of the theorem is now complete.

A problem yet unsolved is the limiting value, if such exists, for $\pi_i(t)$ and $\pi_j(t)$, $j > i$, when $a_i = \infty$. In the proof of Theorem 1 the key-renewal theorem was used to show that

$$\int_0^t \left\{\psi_{i-1}(t-u) - \psi_i(t-u)\right\} dK(u) \to \frac{1}{\mu_1} \int_0^\infty \left\{\psi_{i-1}(u) - \psi_i(u)\right\} du$$

(=0 if $\mu_1 = \infty$), where $K^*(s) = \dfrac{F^*(s)}{1-F^*(s)}$. This result is not necessarily correct if $\psi_{i-1}(u) - \psi_i(u)$ is not Lebesgue-integrable on $(0, \infty)$ i.e. if $a_j \nless \infty$ for all $j \leq i$. The problem is therefore linked to that of finding $\lim_{t\to\infty} \int_0^t q(t-u)\, dK(u)$, and of course the conditions for this limit to exist, where $K^*(s) = \dfrac{F^*(s)}{1-F^*(s)}$, $q(u)$ is a function of bounded variation and $F(t)$ a distribution function with infinite mean.

* * * * * * *

§ 3. Applications to queueing theory.

Throughout this section, Kendall's notation for queueing systems will be employed. Thus, $M/E_k/1$ will denote the single server queue with Poisson input and Erlangian service time distribution. $A(x)$ and $B(x)$ denote the input and service distributions respectively and a_1 and b_1 their respective means. w_n is the time the nth arriving customer waits before his service begins and $W(t)$ the virtual waiting time, viz. the time a customer would have to wait if he arrived at time t. Service is given on a "first-come-first-served" basis. An idle period (i.p.) is defined as the interval between the departure of the last customer in the queue and the first subsequent arrival of a customer, while a busy period (b.p.) is defined as the interval between the end of an idle period and the beginning of the next one. The number of services in a busy period is a random variable, N. Quantities that we are interested in in this section are $E[N]$, $E[b.p.]$, $E[i.p.]$ and $P[W(t) = 0]$.

The theory of queues will be considered in greater detail in Chapter IV and Chapter V.

It will be assumed throughout that $E[b.p.] < \infty$ or, equivalently, $b_1 < a_1$. The equivalence of these two assumptions was proved by Lindley [1952]. It will also be assumed that at $t = 0$, the first customer arrives to be served, and $a_1 < \infty$.

For the general queue $GI/G/1$, we consider the 2-state process $z(t)$ whose states are "server busy" (1), and "server idle"(2).

Then,

$$X_1 \equiv \text{b.p.}, \quad X_2 \equiv \text{i.p.}, \quad \text{and} \quad \pi_2(t) = P[W(t) = 0] \; .$$

From Theorem 1,
$$\lim_{t\to\infty} P\,[W(t) = 0] = \frac{E[X_2]}{E[X_1 + X_2]} \tag{3.1}$$

The sequence of inter-arrival times $\{t_i\}$ forms a stochastic process with finite means a_1, and since $E[X_1] < \infty$, $E[N] < \infty$. Moreover, the event $\{N = n\}$ and the random variables t_{n+1}, t_{n+2}, ... are mutually independent.

$\therefore$ by Wald's theorem (Johnson, 1959)

$$E[X_1 + X_2] = E[t_1 + t_2 + \cdots + t_N]$$

$$= a_1\, E[N] \quad .$$

Similarly,

$$E[X_1] = E[s_1 + s_2 + \cdots + s_N]$$

$$= b_1\, E[N], \quad \text{where}$$

s_i is the service time of the ith customer.

$$\therefore \quad E[X_2] = E[N]\,\{a_1 - b_1\} \quad .$$

From (3.1),

$$\lim_{t\to\infty} P\,[W(t) = 0] = \frac{a_1 - b_1}{a_1}$$

$$= 1 - \rho \quad \text{where} \quad \rho = \frac{b_1}{a_1} \tag{3.2}$$

It will be shown in Chapter IV that if $E[X_1] < \infty$, an appeal to a fundamental result in renewal theory yields

$$\lim_{t\to\infty} P[w_n = 0] = \frac{1}{E[N]} \quad . \tag{3.3}$$

Suppose

$$\lim_{n\to\infty} P\,[w_n = 0] = \lim_{t\to\infty} P\,[W(t) = 0] \quad . \tag{3.4}$$

Then, from (3.3) and (3.2)

$$E[N] = \frac{1}{1-\rho} \ , \quad E[X_1] = \frac{b_1}{1-\rho} \quad \text{and} \quad E[X_2] = \frac{a_1 - b_1}{1-\rho} \ . \tag{3.5}$$

Example 3.1

M/G/1.

Because the input is Poisson, the i.p. and b.p. are independent random variables. For the same reason, the i.p. has the same distribution as the inter-arrival time, $1 - e^{-\lambda x}$, say. Hence, from (2.3)

$$\pi_2^*(s) = \frac{\chi_1^*(s) - \chi_1^*(s)\ \chi_2^*(s)}{1 - \chi_1^*(s)\ \chi_2^*(s)}$$

$$= \frac{\chi_1^*(s)\left\{1 - \frac{\lambda}{\lambda+s}\right\}}{1 - \chi_1^*(s).\ \frac{\lambda}{\lambda+s}}$$

$$= \frac{s\chi_1^*(s)}{\lambda+s - \lambda\chi_1^*(s)}\ ,$$ which agrees with

results of Takács [1962; Theorems 3 and 7 of Chapter 1, Section 3].

From (3.2),

$$\lim_{t\to\infty} P[W(t) = 0] = 1 - \lambda b_1 \left(\text{since } E[X_2] = \frac{1}{\lambda}\right) \tag{3.6}$$

From (3.6) and (3.1),

$$E[X_1] = \frac{b_1}{1 - \lambda b_1} = \frac{b_1}{1-\rho} \quad .$$

However, $E[X_1]$ may be obtained in another way. Takács [1962; Theorem 10, Chapter 1, Section 3] showed that for M/G/1 (3.4) holds. Hence, from (3.5), $E[N] = \frac{1}{1 - \lambda b_1}$,

$$E[X_1] = \frac{b_1}{1-\rho} \quad \text{and} \quad E[X_2] = \frac{\frac{1}{\lambda} - b_1}{1 - \lambda b_1} \quad \frac{1}{\lambda} \quad .$$

Although we have just considered an example of a queue in which the lengths of the i.p. and b.p. are independent, (3.1) and (3.2) may be used to yield results for the more general queue GI/G/1, since our basic result, (ii) of Theorem 1, permits non-independence between distinct states. At the end of § 2, Chapter IV, we derive with the additional aid of fluctuation theory the following expression for $\pi_2^*(s)$, which is true for the queue GI/G/1 :

$$\pi_2^*(s) = 1 - \exp \sum_{k=1}^{\infty} \frac{1}{k} \Big[\int_0^{\infty} e^{-sx} B_n(x)\, dA_n(x) - \int_0^{\infty} e^{-sx} \Big\{ 1 - A_n(x) \Big\}\, dB_n(x$$

Applications of the cyclic-k-state process outside queueing theory, for k = 2, are found in Cox (1962).

For k > 2, the process may be used as a model of the single server system with a Poisson input and a special Erlangian service time distribution, in which no queues are ~~all owed~~ allowed. (A customer arriving while the server is busy leaves the system). Such a queue is discussed by Cox and Smith (1961, Section 5.2). Here, $A(x) = 1 - e^{-\lambda x}$,

$B(x) = 1 - e^{-k\mu t} \sum_{n=0}^{k-1} \frac{(k\mu t)^n}{n!}$. The service time of a

customer may be imagined as being divided into k phases in series. After service in phase 1, the customer proceeds to phase 2 to be served, and so on until he arrives at phase k. On completion of service in the kth phase, the server becomes idle until the arrival of the next customer. If each of the k phases is given a service time distribution $1 - k\mu e^{-k\mu x}$, then it is readily shown that the service time of a customer has distribution $B(x) = 1 - e^{-k\mu t} \sum_{0}^{k-1} \frac{(k\mu t)^n}{n!}$. Now consider the cyclic-k+1-state process $z(t)$ in which i ($i = 1, 2, \ldots, k$) denotes the state "server busy; service in ith phase" and $k+1$ the state "server idle". Then $P\,[W(t) = 0] = \pi_{k+1}(t)$ and by Theorem 1,

$$\lim_{t\to\infty} P\,[W(t) = 0] = \frac{a_{k+1}}{\mu_1} = \frac{E[i.p.]}{E[i.p. + b.p.]} \qquad \text{(cf 3.1)}$$

$$= \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + \frac{1}{\mu}} = \frac{1}{1+\rho} .$$

(Admittedly, the above result is more easily obtained by considering only the two states 'server busy' and 'server idle').

CHAPTER III

Fluctuation Theory and One-Dimensional Random Walk

§ 1. Preliminary Remarks

This is a purely expository chapter, devoted mainly to reviewing briefly the results of fluctuation theory and showing how these results were recently obtaind by Spitzer (1964), by simple analytic arguments applied to the one-dimensional random walk.

Suppose $\{X_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed random variables and that their partial sums are defined by $S_o = 0$, $S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$. Fluctuation theory is concerned with the effect which the fluctuation of these partial sums has on various quantities of interest. Some such typical quantities are :

(a) $T_n^+ = \max_{o \leq k \leq n} \{k / S_j \leq S_k, \ j \leq n\}$, the index of the last maximum in the sequence $\{S_i\}_{i=o}^{n}$;

(b) $T_n = \min_{o \leq k \leq n} \{k / S_j \leq S_k, \ j \leq n\}$, the index of the first maximum;

(c) N_n^+, the number of strictly positive sums in $\{S_i\}_{i=1}^{n}$;

(d) N_n, the number of non-negative sums in $\{S_i\}_{i=1}^n$ $(N_o = 0)$;

and (e) $M_n = S_{T_n}$, the value of the maximum of the first n sums.

The important results of fluctuation theory may be considered as belonging to two major classes. In the first class are the exponential identities which give the distributions of T_n, M_n, etc. in terms of the individual distributions of $S_o, S_1, \ldots, S_n$.

An example of this is the main theorem of Spitzer, (1956),

$$\sum_{n=o}^{\infty} t^n E\left[e^{i\beta S_n + i\theta M_n}\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} \left\{E\left[e^{i(\theta+\beta)S_k};\ S_k \geq 0\right] + \right.$$

$$\left. + E\left[e^{i\beta S_k};\ S_k < 0\right]\right\} \quad ,$$

where

$$E\left[e^{i\beta S_k};\ A\right] = \int_A e^{i\beta x}\, dP\left[S_k \leq x\right] \qquad \text{by definition.}$$

In the other class are the extremal factorizations, of which Andersen's (1953) $P[N_n = k] = P[N_k = k]\ P[N_{n-k} = 0]$ provides an example, which enable the distribution of a random variable to be determined from its extremal values.

In § 2 of this chapter, we shall exemplify the combinatorial approach to fluctuation problems by following Port (1963) in using Feller's equivalence principle to arrive at the familiar results of Andersen, Spitzer, Baxter, et al.

§ 3 will be concerned with the other major method used in fluctuation theory, the analytic method. Discussion will be focussed on Chapter IV of Spitzer's book (1964). In this chapter of his book, Spitzer solved the major problems of fluctuation theory by considering random walk on the half-line and appealing to theorems in complex variable analysis.

Of key importance in both the analytic and combinatorial approaches is the fact that the variable T_n is a ladder index, whence the occurrence of a maximum is a recurrent event.

k is called a ladder index for the sequence $\{S_i\}_{i=0}^{n}$ if $S_j \leq S_k$ for $0 \leq j \leq k-1$. Clearly, $\ell > k$ is a ladder index if and only if $S_\ell \geq S_j$ for $j = k, \ldots, \ell-1$ i.e. if and only if $S_\ell - S_k \geq S_j - S_k$, $j = k, \ldots, \ell-1$. As the last inequality is a function of only $X_{k+1}, \ldots, X_\ell$, the occurrence of ladder indices of $\{S_i\}$ constitutes a recurrent event.

It is possible to go even further. Port (1963, Theorem III, 3) was able to show, with little difficulty, that if $\{W_k\}$ and $\{Z_k\}$ are sequences of ladder indices (the index 0 is excluded), and lengths of intervals between successive indices respectively, then

(Z_k, W_k) are independently and identically distributed.

Thus,

$$E\left[e^{i\theta(Z_1 + \cdots + Z_k)} t^{W_1 + \cdots + W_k}\right] = \left\{E\left[e^{i\theta Z_1} t^{W_1}\right]\right\}^k .$$

We end these preliminary remarks with the following definitions :

(i) if X is a random variable and A an event, then
$$E[f(X);\, A] = \int_A f(x)\, dP[X \leq x] \quad ;$$

(ii) $$E[f(X)/A] = \int_{-\infty}^{\infty} f(x)\, dP[X \leq x/A] \quad ;$$

(iii) $$\phi(\theta) = E\left[e^{i\theta X_k}\right] .$$

(iv) $S_k^+ = \max(0, S_k)$; and

(v) $S_k^- = \max(0, -S_k)$.

* * * * * * * *

§ 2. The Combinatorial Method.

An example of the use of a combinatorial result in proving fluctuation theorems is to be found in Feller's Equivalence Principle (1959). This principle, first obtained by Andersen (1953) by combinatorial arguments, asserts the stochastic equivalence of T_n and N_n^+, T_n^+ and N_n, T_{no} and N_n^-, and $\bar{T}_{no}$ and $\bar{N}_n$, where T_{no} is the index of the first minimum sum in the sequence $\{S_k\}_{k=o}^{n}$, $\bar{T}_{no}$ the index of the last minimum; N_n^- the number of strictly negative sums in the sequence $\{S_k\}_{k=1}^{n}$, and $\bar{N}_n$ the number of nonpositive ones; and T_n, T_n^+, N_n and N_n^+ are defined as in §1. More precisely, the

Equivalence Principle states that for $j = 0, 1, \ldots, n$

$$(2.1) \qquad P[T_n = j] = P[N_n^+ = j],$$

$$(2.2) \qquad P[T_n^+ = j] = P[N_n = j],$$

$$(2.3) \qquad P[T_{no} = j] = P[N_n^- = j] \qquad \text{and}$$

$$(2.4) \qquad P[\bar{T}_{no} = j] = P[\bar{N}_n = j] \quad .$$

Briefly, Feller's proof of the principle proceeds as follows. Let $(Y_1, \ldots, Y_n)$ be the vector obtaind by applying some permutation σ on the indices of $(X_1, \ldots, X_n)$ such that $Y_j = X_{\sigma(j)}$. The principle is trivially true for $n = 1$.

Suppose it is true for $n - 1$ and consider the case $S_n = X_1 + \dots + X_n = Y_1 + \dots + Y_n \leq 0$. Since T_n cannot be n in this case, the induction hypothesis gives

$$P\left[T_n = j \middle/ Y_n = X_\nu\right] = P\left[N_n^+ = j \middle/ Y_n = X_\nu\right] \quad \text{for } j = 1, \dots, n-1$$

and $\nu = 1, \dots, n$.

(2.1) now follows if the last equation is multiplied by $\frac{1}{n}$ and summed over ν, since obviously $P[Y_n = X_\nu] = \frac{1}{n}$. For the case $S_n > 0$, the same argument yields (2.3). As (2.1) and (2.3) are equivalent, their proof is now complete. (2.2) and (2.4) may be proved similarly.

We now proceed to prove Spitzer's basic identity (1956, Th. 6.1) by using only the equivalence principle and the fact that ladder indices are recurrent events.

Since T_n^+ is a ladder index,

$$\{T_n^+ = k\} = \{T_k^+ = k\} \cap \{T_{n-k}^+ = 0\} \quad \text{and}$$

$$E\left[e^{i\theta S_n}; T_n^+ = k\right] = E\left[e^{i\theta S_k} e^{i\theta(S_n - S_k)}; T_n^+ = k\right]$$

$$= E\left[e^{i\theta S_k}; T_k^+ = k\right] E\left[e^{i\theta S_{n-k}}; T_{n-k}^+ = 0\right] .$$

Multiplying both sides by $x^k t^n$ and summing over $0 \leq k \leq n < \infty$, we get

$$\sum_{n=0}^{\infty} E\left[e^{i\theta S_n} x^{T_n^+}\right] t^n = \sum_{n=o}^{\infty} E\left[e^{i\theta S_n};\ T_n^+ = n\right] (xt)^n \sum_{n=o}^{\infty} E\left[e^{i\theta S_n}; T_n^+ = 0\right] t^n \tag{2.5}$$

Spitzer's basic identity is, with obvious modification of notation,

$$\sum_{n=o}^{\infty} E\left[e^{i\theta S_n + i\gamma M_n}\right] t^n$$

$$= \exp \sum_{k=1}^{\infty} \frac{t^k}{k} \left\{E\left[e^{i(\theta+\gamma)S_k};\ S_k \geq 0\right] + E\left[e^{i\gamma S_k};\ S_k < 0\right]\right\} \tag{2.6}$$

(2.6) is easily shown (Port, 1963) to be equivalent to each of the two following identities

$$\sum_{n=o}^{\infty} E\left[e^{i\theta S_n};\ T_n^+ = n\right] t^n = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k};\ S_k \geq 0\right] \tag{2.7}$$

$$\sum_{n=o}^{\infty} E\left[e^{i\theta S_n};\ T_n^+ = 0\right] t^n = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k};\ S_k < 0\right] \tag{2.8}$$

That (2.7) and (2.8) are equivalent may be proved with the aid of (2.5). Setting x = 1 in (2.5)

$$\frac{1}{1-t\phi(\theta)} = \sum_{n=o}^{\infty} E\left[e^{i\theta S_n}\right] t^n = \sum_{n=o}^{\infty} E\left[e^{i\theta S_n}; T_n^+=n\right] t^n \sum_{n=o}^{\infty} E\left[e^{i\theta S_n}; T_n^+=0\right] t^n \tag{2.9}$$

But

$$\frac{1}{1-t\phi(\theta)} = \exp \sum_{n=1}^{\infty} \frac{t^n}{n} E\left[e^{i\theta S_n}\right]$$

$$= \exp \sum_{n=1}^{\infty} \frac{t^n}{n} \left\{ E\left[e^{i\theta S_n};\ S_n \geq 0\right] + E\left[e^{i\theta S_n};\ S_n < 0\right] \right\} .$$

The last equation and (2.9) show that (2.7) and (2.8) are equivalent.

By (2.9) and (2.5)

$$\sum_{n=0}^{\infty} E\left[e^{i\theta S_n} x^{T_n^+}\right] t^n = \frac{P(\theta, xt)}{\{1-t\phi(\theta)\}\ P(\theta,t)} \tag{2.10}$$

where

$$P(\theta,t) = \sum_{n=0}^{\infty} E\left[e^{i\theta S_n};\ T_n^+ = n\right] t^n .$$

Differentiating (2.10) with respect to x at $x = 1$,

$$\sum_{n=0}^{\infty} E\left[e^{i\theta S_n} T_n^+\right] t^n = \frac{t\ P'(\theta,t)}{\{1-t\phi(\theta)\}\ P(\theta,t)} .$$

Hence

$$\frac{P'(\theta,t)}{P(\theta,t)} = \sum_{n=1}^{\infty} t^{n-1} E\left[e^{i\theta S_n} \left\{T_n^+ - T_{n-1}^+\right\}\right] .$$

(We have used the fact that the X_i's are independently and identically distributed to deduce

$$E\left[e^{i\theta X_{n+1}}\right] E\left[e^{i\theta S_n} T_n^+\right] = E\left[e^{i\theta X_{n+1}} \Big/ T_n^+ = k\right] E\left[e^{i\theta S_n} \Big/ T_n^+ = k\right] E[T_n^+]$$

$$= E\left[e^{i\theta S_{n+1}} T_n^+\right] . \quad)$$

As $T_o^+ = 0$ by definition, $P(\theta, 0) = 1$ and by solving the differential equation,

$$\sum_{n=o}^{\infty} E\left[e^{i\theta S_n}; T_n^+ = n\right] t^n = \exp \sum_{n=1}^{\infty} \frac{t^n}{n} E\left[e^{i\theta S_n} (T_n^+ - T_{n-1}^+)\right] . \tag{2.11}$$

We use now, for the first time, the equivalence principle. By this principle,

$$E\left[e^{i\theta S_n} (T_n^+ - T_{n-1}^+)\right] = E\left[e^{i\theta S_n} (N_n - N_{n-1})\right] .$$

But N_n is N_{n-1} or $(N_{n-1} + 1)$ according as whether $S_n < 0$ or $S_n \geq 0$. Hence

$$E\left[e^{i\theta S_n} (T_n^+ - T_{n-1}^+)\right] = E\left[e^{i\theta S_n}; S_n \geq 0\right] ,$$ and (2.7) now follows from (2.11).

We have thus proved Spitzer's basic identity. In the process, no more elaborate mathematical tool was needed than the equivalence principle and the recurrence property of ladder indices.

The rest of this section will show how, without the aid of any new method, Spitzer's basic identity yields most of the familiar results of fluctuation theory. But first, it is necessary to derive by simple probabilistic arguments and the fact that $\{(Z_k, W_k)\}$ are independent and identically distributed, two additional identities. $\big((Z_k, W_k)$ was defined at the end of § 1.$\big)$

$$\frac{1}{1-E\left[e^{i\theta Z_1} t^{W_1}\right]} = \sum_{n=o}^{\infty} t^n E\left[e^{i\theta S_n};\ T_n^+ = n\right] \tag{2.12}$$

$$\frac{1-E\left[e^{i\theta Z_1} t^{W_1}\right]}{1 - t\phi(\theta)} = \sum_{n=o}^{\infty} t^n E\left[e^{i\theta S_n};\ T_n^+ = 0\right] \tag{2.13}$$

The proof of (2.12) (see Port, 1963) relies mainly on the fact that

$$E\left[e^{i\theta(Z_1+ \cdots + Z_k)} t^{W_1 + \cdots + W_k}\right] = \left\{E\left[e^{i\theta Z_1} t^{W_1}\right]\right\}^k,$$

whence

$$\frac{1}{1-E\left[e^{i\theta Z_1} t^{W_1}\right]} = 1 + \sum_{k=1}^{\infty} E\left[e^{i\theta(Z_1+\cdots+Z_k)} t^{W_1+\cdots+W_k}\right]$$

$$= 1 + \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} t^n E\left[e^{i\theta S_n}; W_1+ \cdots + W_k = n\right].$$

(2.12) now follows from the equivalence $\bigcup_{k=1}^{\infty} \{W_1 + \cdots + W_k = n\} = \{T_n^+ = n\}$.

To prove (2.13), the equivalence $\{T_n^+ = 0\} = \{W_1 > n\}$ is used. We have

$$\{1-t\phi(\theta)\} \sum_{n=0}^{\infty} t^n E\left[e^{i\theta S_n};\ T_n^+=0\right] = 1 + \sum_{n=1}^{\infty} t^n \left\{E\left[e^{i\theta S_n};\ T_n^+=0\right] - E\left[e^{i\theta S_n};\ T_{n-1}^+ = 0\right]\right\}$$

$$= 1 - \sum_{n=1}^{\infty} t^n E\left[e^{i\theta S_n};\ W_1 = n\right]$$

$$= 1 - E\left[e^{i\theta Z_1} t^{W_1}\right] .$$

Using the recurrence property of ladder indices, the basic identities (2.7) and (2.8) may be expressed in the form (2.6) in which they first appeared in Spitzer's paper (1956).

$$E\left[e^{i\theta S_n + i\gamma M_n};\ T_n^+ = k\right] = E\left[e^{i(\theta+\gamma)S_k + i\theta(S_n - S_k)};\ T_n^+ = k\right]$$

$$= E\left[e^{i(\theta+\gamma)S_k};\ T_k^+ = k\right] E\left[e^{i\theta(S_n-S_k)};T_{n-k}^+=0\right].$$

Multiplying both sides of the equation by $x^k t^n$ and summing over $0 \leq k \leq n < \infty$,

$$\sum_{0}^{\infty} t^n \, E\left[e^{i\theta S_n + i\gamma M_n} x^{T_n^+}\right] = \sum_{n=0}^{\infty} E\left[e^{i(\theta+\gamma)S_n}; T_n^+ = n\right](xt)^n \sum_{n=0}^{\infty} E\left[e^{i\theta S_n}; T_n^+ = 0\right] t^n$$

$$= \exp \sum_{k=1}^{\infty} \frac{(xt)^k}{k} E\left[e^{i(\theta+\gamma)S_k}; S_k \geq 0\right]$$

$$\exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k}; S_k < 0\right]$$

by (2.7) and (2.8) .

(2.6) is the special case $x = 1$.

Putting $\theta = 0$ in (2.7) and (2.8), we get the two following identities of Andersen (1953) and Feller (1959) :

$$\sum_{n=0}^{\infty} t^n \, P\left[T_n^+ = n\right] = \exp \sum_{t=1}^{\infty} \frac{t^k}{k} P\left[S_k \geq 0\right] \qquad (2.14)$$

$$\sum_{n=0}^{\infty} t^n \, P\left[T_n^+ = 0\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} P\left[S_k < 0\right] \qquad (2.15)$$

Taking reciprocals in (2.12) and then applying (2.7) yields ~~this~~ the following result of Baxter and Spitzer,

$$E\left[e^{i\theta Z_1} t^{W_1}\right] = 1 - \exp\left\{-\sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k};\ S_k \geq 0\right]\right\} \qquad (2.16)$$

From the equivalence principle, (2.5), (2.7) and (2.8),

$$\sum_{n=0}^{\infty} E\left[e^{i\theta S_n} x^{N_n}\right] t^n = \exp \sum_{n=1}^{\infty} \frac{(xt)^n}{n} E\left[e^{i\theta S_n}; S_n \geq 0\right] \exp \sum_{n=1}^{\infty} \frac{t^n}{n} E\left[e^{i\theta S_n}; S_n < 0\right]$$

(2.17)

(2.17) is a result of Baxter and, for $\theta = 0$, of Andersen.

The manner of approach adopted in this section, that of proving Spitzer's basic identity ((2.7) and (2.8)) from Feller's equivalence principle and the recurrence property of ladder indices, and subsequently using this identity to derive the other familiar results, is due to Port (1963) and Feller (1959).

* * * * * * *

§ 3. One-dimensional random walk

In the preceding chapter, identities were established which related the various variables M_n, T_n^+, N_n to the partial sums S_k of the random variables X_1, X_2, ..., which could be discrete or continuous. This was done by resorting to combinatorial arguments. By considering the one-dimensional random walk with discrete variables X_i, Spitzer (1964) was able to arrive at the results of § 2 by simple analytic methods. His basic identity (3.2) was proved by appealing only to theorems of Morera and Liouville in complex variable analysis. Other proofs made use of the fact that occurrences of ladder indices are recurrent events. The recurrence property of ladder indices consistently simplifies what would otherwise be extremely difficult problems and its importance cannot be over-emphasised. That it is possible to get results from only a knowledge of the individual distributions of the partial sums S_1, ..., S_k may in fact be due to this property. We recall that this recurrence property played a prominent part in the combinatorial method as well.

A major part of this section will be concerned with the results and methods of proofs of Chapter 4 of Spitzer's (1964) book. Except where specifically stated otherwise, all references made to Spitzer will be to this particular chapter of his book.

Spitzer considered only the case in which the steps of the walk assume integral values, but as will be indicated, all his results with which we are concerned in this section hold for the continuous case as well.

The problems of fluctuation theory may be formulated in terms of the random walk $\underset{\sim}{x}_n$ which begins at $\underset{\sim}{x}_o = 0$ and after n "steps", reaches $\underset{\sim}{x}_n = X_1 + X_2 + \dots + X_n$, where each X_i, the length of the ith "step", is an integer-valued random variable that is independently and identically distributed with $P[X_i \leq x] = P[0,x]$. The degenerate case $P(0,0) = 1$ is excluded. Let $S_k = X_1 + \dots + X_k$, $S_o = 0$. We are now interested in the variables

$$M_n = \max_{0 \leq k \leq n} S_k ,$$

$$T_n = \min_{0 \leq k \leq n} [k/S_k = M_n] \qquad \text{and}$$

N_n^+, the number of positive terms in the sequence $\{S_i\}_{i=o}^{n}$.

By Fourier analysis and complex variable theory, Spitzer obtained expressions identical with those in § 2. Spitzer also obtained identities for two other variables of interest, the "hitting times" on the half-lines $[1,\infty)$ and $[0,\infty)$, defined by

$$T = \min_{1 \leq n \leq \infty} [n/S_n > 0]$$

$$T' = \min_{1 \leq n \leq \infty} [n/S_n \geq 0] \quad .$$

Note that T and T', like T_n, are ladder indices. Further T' is W_1 and $S_{T'}$ is Z_1, where W_1 and Z_1 are the random variables defined in § 1, i.e. T' is the 1st ladder index.

The concept of the reversed random walk $\underset{\sim}{x}_n^R$ is a convenient one which Spitzer used frequently and effectively. For $\underset{\sim}{x}_n^R$, $P(0,x) = P(0, -x)$ and $S_k^R = -S_k$. (The notation here is obvious). Define $T^* = \min_{1 \leq n \leq \infty} [n/S_n < 0]$ and $T^{*\prime} = \min_{1 \leq n \leq \infty} [n/S_n \leq 0]$. Then clearly T^* and $T^{*\prime}$ play the roles of T and T' respectively in the reversed random walk, $\underset{\sim}{x}_n^R$

The main identities established by Spitzer are the following :

$$1 - E\left[t^T \xi^{S_T}\right] = \exp\left\{-\sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k}; S_k > 0\right]\right\} \quad |\xi| \leq 1 \qquad (3.1)$$

$$\sum_{k=0}^{\infty} t^k E\left[\xi^{S_k}; T > k\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k}; S_k \leq 0\right] \quad |\xi| \geq 1 \qquad (3.2)$$

$$\sum_{k=0}^{\infty} t^k E\left[\xi^{M_k} \eta^{M_k - S_k}\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} \left\{ E\left[\xi^{S_k}; S_k > 0\right] + E\left[\eta^{-S_k}; S_k \leq 0\right] \right\}$$

$$|\xi| \leq 1, \quad |\eta| \leq 1 \qquad (3.3)$$

$$P\left[N_n^+ = k\right] = P\left[N_k^+ = k\right] P\left[N_{n-k}^+ = 0\right] \qquad 0 \leq k \leq n \qquad (3.4)$$

$$\sum_{k=0}^{\infty} t^k P\left[N_k^+ = 0\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} P\left[S_k \leq 0\right] \qquad (3.5)$$

$$\sum_{k=0}^{\infty} t^k P\left[N_k^+ = n\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} P\left[S_k > 0\right] \qquad (3.6)$$

where $0 \leq t < 1$, and ξ and η are complex numbers. For the other hitting time T', we have

$$1 - E\left[t^{T'} \xi^{S_{T'}}\right] = \exp\left\{-\sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k};\ S_k \geq 0\right]\right\} \qquad (3.1')$$

and

$$\sum_{k=0}^{\infty} t^k E\left[\xi^{S_k};\ T' > k\right] = \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k};\ S_k < 0\right] \qquad (3.2')$$

Note that as $\{T' > k\} = \{T_k^+ = 0\}$, (3.2') is simply Spitzer's basic identity, (2.8) of § 2, with $\xi = e^{i\theta}$. Also, (3.1') is (2.16) of § 2 and (3.3) is the basic identity in another form (c.f. (2.6) of § 2).

Just as it was possible in § 2 to obtain (2.6) from (2.7) and (2.8) using the recurrence property of the ladder index T_n^+, we can derive (3.3) from (3.2) and (3.2'); the

ladder index in this case is the variable T_n. The derivation of (3.3) will serve to illustrate the importance of the recurrent nature of ladder indices in the walk, and Spitzer's use of the concept of the reversed walk.

The crucial step in the derivation is the argument which yields

$$E\left[\xi^{S_k}\eta^{S_k-S_n};\ S_j < S_k,\ 0 \le j < k;\ S_j \le S_k,\ k < j \le n\right]$$

$$= E\left[\xi^{S_k};S_j<S_k,\ 0\le j<k\right]\ E\left[\eta^{S_k-S_n};S_j\le S_k,\ k<j\le n\right] \qquad (3.7)$$

(3.7) holds because, k being the index of the first maximum of the partial sums $\{S_i\}_0^{n\ge k}$, and hence a ladder index, the parts within the two parentheses on the right hand side are mutually independent. Since

$$\left\{S_j<S_k;\ 0\le j<k\right\} = \left\{\sum_{i=j+1}^{k} X_i > 0;\ 0\le j<k\right\} = \left\{S_j > 0;\ 0<j\le k\right\} \quad \text{and}$$

$$\left\{S_j \le S_k;\ k<j\le n\right\} = \left\{\sum_{i=k+1}^{j} X_i \le 0;\ k<j\le n\right\} = \left\{S_j \le 0;\ 0<j\le n-k\right\},$$

by the definition of T and T*', it follows from (3.7) that

$$\sum_{n=0}^{\infty} t^n E\left[\xi^{M_n} \eta^{M_n - S_n}\right] = \sum_{n=0}^{\infty} t^n \sum_{k=0}^{n} E\left[\xi^{S_k} \eta^{S_k - S_n};\ T_n = k\right]$$

$$= \sum_{n=0}^{\infty} t^n \sum_{k=0}^{n} E\left[\xi^{S_k};\ T^{*\prime} > k\right] E\left[\eta^{-S_{n-k}};\ T > n-k\right]$$

$$= \sum_{n=0}^{\infty} t^n E\left[\xi^{S_n};\ T^{*\prime} > n\right] \sum_{n=0}^{\infty} t^n E\left[\eta^{-S_n};\ T > n\right] . \tag{3.8}$$

The concept of the reversed random walk $\underset{\sim}{x}_n^R$ is now introduced. Since $P^R(0,x) = P(0,-x)$, $S_k^R = -S_k$ and $T^{*\prime} = T'^R$.

Hence

$$\sum_{n=0}^{\infty} t^n E\left[\xi^{S_n};\ T^{*\prime} > n\right] = \sum_{n=0}^{\infty} t^n E\left[\xi^{-S_n^R};\ T'^R > n\right]$$

$$= \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{-S_n^R};\ S_k^R < 0\right] \qquad \text{by } (3.2')$$

$$= \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_n};\ S_k > 0\right] .$$

This last result, with (3.8) and (3.2), yields (3.3).

In proving (3.1) and (3.2), Spitzer had to rely on the recurrence property of the hitting time T. However, the use of this property at this early stage is unnecessary, and we give here a modified form of his proof.

A major step in the proof is the establishment of the identity

$$1 - E\left[t^T e^{i\theta S_T}\right] = \left\{1 - t\phi(\theta)\right\} \sum_{k=o}^{\infty} t^k E\left[e^{i\theta S_k}; T > k\right] \qquad (3.9)$$

(c.f. (2.13)), where $\phi(\theta) = E[e^{i\theta X_1}]$. It is at this step that Spitzer appeals to the recurrence property of T. Proving (3.9) in a different manner, we observe that

$$\left\{1-t\phi(\theta)\right\} \sum_{n=o}^{\infty} t^n E\left[e^{i\theta S_n}; T>n\right] = 1 + \sum_{n=1}^{\infty} E\left[e^{i\theta S_n}; T>n\right] t^n$$

$$- \sum_{n=1}^{\infty} E\left[e^{i\theta S_n}; T>n-1\right] t^n$$

$$= 1 - \sum_{n=1}^{\infty} E\left[e^{i\theta S_n}; T = n\right] t^n$$

$$= 1 - E\left[t^T e^{i\theta S_T}\right] .$$

Now,

$$1 - t\phi(\theta) = \exp\left\{- \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k}\right]\right\}$$

$$= \exp\left\{- \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k}; S_k>0\right]\right\} \exp\left\{- \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k}; S_k\leq 0\right]\right\}$$

$$(3.10)$$

From (3.9) and (3.10),

$$\left\{1 - E\left[t^T e^{i\theta S_T}\right]\right\} \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k}; S_k > 0\right] =$$

$$= \sum_{k=0}^{\infty} t^k E\left[e^{i\theta S_k}; T > k\right] \exp\left\{- \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[e^{i\theta S_k}; S_k \leq 0\right]\right. \tag{3.11}$$

At this stage, Spitzer embarked on the analytic arguments that characterize this type of proof. It is easily verified that $1 - E\left[t^T \xi^{S_T}\right]$ and $\exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k}; S_k > 0\right]$ are analytic in $|\xi| < 1$ and continuous on $|\xi| \leq 1$. Similarly, $\sum_{k=0}^{\infty} t^k E\left[\xi^{S_k}; T > k\right]$ and $\exp\left\{- \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k}; S_k \leq 0\right]\right\}$ are analytic in $|\xi| > 1$ and continuous and bounded in $|\xi| \geq 1$. Thus, from (3.11) and the theorems of Morera and Liouville (Titchmarsh, Theory of Functions, 1939), there exists a constant K such that

$$\left\{1 - E\left[t^T \xi^{S_T}\right]\right\} \exp \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k}; S_k > 0\right] = K \quad \text{for} \quad |\xi| < 1$$

and

$$\sum_{k=0}^{\infty} t^k E\left[\xi^{S_k}; T > k\right] \exp\left\{- \sum_{k=1}^{\infty} \frac{t^k}{k} E\left[\xi^{S_k}; S_k \leq 0\right]\right. = K$$

for $|\xi| > 1$.

Letting $\xi \to 0$ in one of the above equations, we get $k = 1$. (3.1) and (3.2) now follow.

(3.1') and (3.2') may be proved similarly.

It is clear, from the definitions of N_n^+, T and T*',

that $\{N_n^+ = 0\} = \{T > n\}$ and $\{N_n^+ = n\} = \{T^{*\prime} > n\}$. With these equivalences, putting $\xi = 1$ in (3.1) and (3.1') yields (3.5) and (3.6). To get from (3.1') to (3.6), Spitzer employed the useful device of the reversed walk again.

(3.4) is the well-known extremal factorization of Andersen (1953). To prove it, Spitzer resorts to Fourier analysis and complex variable theory. In fact, Spitzer proves a more general result,

$$\sum_{n=0}^{\infty} t^n E\left[e^{i\theta S_n} x^{N_n^+}\right] = \exp \sum_{k=1}^{\infty} \left\{\frac{t^k}{k} E\left[e^{i\theta S_k}; S_k \le 0\right] + \frac{(xt)^k}{k} E\left[e^{i\theta S_k}; S_k > 0\right]\right\} \tag{3.12}$$

(cf. (2.17) of § 2). (3.4) is easily deduced from (3.12), (3.5) and (3.6).

Excepting (3.12) and (3.4), the discreteness of the random variables X_i has not been used in any of the proofs so far; all the identities, with the possible exceptions of (3.12) and (3.4), are therefore equally valid for continuous variables X_i. That (3.12) and (3.4) remain true even when the X_i are continuously distributed will be shown at the end of this section.

The characteristic feature of the exponential identities (3.1), (3.1'), (3.2), (3.2'), (3.3) and (3.12), indeed the characteristic feature of all these exponential Spitzer-type identities, is that the right hand sides of the equalities involve only the characteristic function of the partial sums S_k^+ or S_k^-, so that a knowledge of the S_k - distribution will, in theory at least, suffice to determine the distributions of other variables of interest. Thus in (3.1), (3.3) and (3.12), we are able to obtain the joint characteristic functions of T and S_T, M_n and S_n, and S_n and N_n^+ respectively in terms of the individual (NOT joint) distributions of S_1, S_2, ...

The identities (3.1) - (3.6) and (3.12), though important, do not answer all the questions regarding the one-dimensional random walk. One would like to know the various conditions under which T, E[T] and E[S_T] are finite and their finite values. Similarly, one is interested in knowing when M_n, T_n and N_n^+ tend in distribution to finite variables M, $\mathcal{J}$ and N , and the values of these limits. With the aid of the identities, Spitzer showed that the finiteness of the quantities in question depends on the finiteness of the series $\sum_{k=1}^{\infty} \frac{1}{k} P[S_k > 0]$. He produces an even simpler criterion by showing that if $m = \sum_{x=-\infty}^{\infty} |x| \; P[X_i = x]$ exists, then $\sum_{k=1}^{\infty} \frac{1}{k} P[S_k > 0]$ is divergent if and only if the mean length of a "step" in the walk, $\nu = \sum_{x=-\infty}^{\infty} x\, P[X_i = x]$ is non-negative.

Spitzer first showed that $P[T < \infty] = 1$ if and only if $\sum_{k=1}^{\infty} \frac{1}{k} P[S_k > 0] = \infty$. This was easily done by putting $\xi = 1$ in (3.1) and letting $t \to 1-$. Furthermore, if $\sum_{x=-\infty}^{\infty} |x|\, P[X_i = x] < \infty$, then $\sum_{k=1}^{\infty} \frac{1}{k} P[S_k > 0] = \infty$ if and only if $\nu = \sum_{x=-\infty}^{\infty} x\, P[X_i = x] \geq 0$.

If $\nu > 0$, then $\frac{S_n}{n} \to \nu > 0$ with probability one and hence $T < \infty$ with probability one. $\nu = 0$ implies that the random walk is recurrent (even if the X_i's are continuous), hence that $T < \infty$. Finally, if $\nu < 0$, an application of the strong law of large numbers shows that $S_n \to -\infty$, so that T cannot be finite with probability one.

A similar result obviously holds with T' and $P[S_k \geq 0]$ replacing T and $P[S_k > 0]$.

An interesting duality between T and the variables M_n, T_n and N_n^+ exist in the following sense. While T is finite with probability one if and only if $\sum_{k=1}^{\infty} \frac{1}{k} P[S_k > 0] = \infty$, M_n, T_n and N_n^+ tend in distribution to finite variables M, $\mathcal{J}$ and N^+ respectively if and only if $\sum_{k=1}^{\infty} \frac{1}{k} P[S_k > 0] < \infty$ (or equivalently, if and only if $\nu < 0$ in the case $m < \infty$). The proof of this appears in detail in Spitzer's classic paper of 1956 (Th. 4.1 and Th. 5.2).

When $M < \infty$, by multiplying (3.3) by (1-t), putting $\eta = 1$ and finally applying an abelian argument, we get

$$E\left[\xi^M\right] = \exp\left\{- \sum_{k=1}^{\infty} \frac{1}{k} E\left[1 - \xi^{S_k};\ S_k > 0\right]\right\}, \qquad |\xi| \leq 1 .$$

By a similar use of (3.12), the corresponding result for N^+ is obtained.

From now on, it will be assumed that $m < \infty$. Thus

$$\sum_{k\ 1}^{\infty} \frac{1}{k} \ P\left[S_k > 0\right] = \infty \text{ if and only if } \nu \geq 0.$$

When T is an honest random variable, i.e. when $\nu \geq 0$, the means E[T] and $E[S_T]$ are of interest. Using the equation

$$E[T] = \lim_{t \to 1-} \frac{1-E[t^T]}{1-t} ,$$

we see that (3.1) yields

$$E[T] = \exp \sum_{k=1}^{\infty} \frac{1}{k} P\left[S_k \leq 0\right] \leq \infty .$$

It was shown earlier that $\sum_{k=1}^{\infty} \frac{1}{k} P\left[S_k \geq 0\right] = \infty$ if and only if $\nu \geq 0$. Applying this to the reversed walk, we get

$$\sum_{k=1}^{\infty} \frac{1}{k} P\left[S_k \leq 0\right] = \infty \text{ if } \nu = 0, \quad \text{and} \quad \sum_{k=1}^{\infty} \frac{1}{k} P\left[S_k \leq 0\right] < \infty \text{ if } \nu > 0.$$

If $\nu > 0$, Wald's lemma in sequential analysis (N.L. Johnson, 1959) yields $E[S_T] = \nu E[T]$.

We pause here to summarize Spitzer's results.

Theorem 1

(a) $\lim_{n\to\infty} M_n = M < \infty$ with probability one if and only if $\nu < 0$.

(b) $\lim_{n\to\infty} N_n^+ = N < \infty$ " " " "

(c) $\lim_{n\to\infty} T_n = \mathcal{T} < \infty$ " " " "

(d) If $\nu < 0$, then

$$E[\xi^M] = \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} E\left[1-\xi^{S_k};\ S_k > 0\right]\right\} \quad |\xi| \leq 1 \quad .$$

$$E[t^{N+}] = E[t^{\mathcal{T}}] = \exp \sum_{k=1}^{\infty} \frac{(t^k-1)}{k} P[S_k > 0] \quad |t| \leq 1.$$

Theorem 2 $\quad P[T < \infty] = 1$ if and only if $\nu \geq 0$.

Theorem 3 $\quad$ If $\nu > 0$, then $E[T] = \exp \sum_{k=1}^{\infty} \frac{1}{k} P[S_k \leq 0] < \infty$ and

$$E[S_T] = \nu E[T].$$

Theorem 4 $\quad$ If $\nu = 0$, then $E[T] = \infty$.

The identical result for T_n and N_n^+ is hardly surprising when we recall that the two variables are stochastically equivalent (equivalence principle, § 2).

Spitzer proved an additional result which he later used to derive the arc-sine law for the special case $\nu = 0$ and $\sigma^2 = \sum_{x=-\infty}^{\infty} x^2 P[X_i = x] < \infty$. Omitting the proof, we state this result as

Theorem 5
If $\nu = 0$ and $\sigma^2 < \infty$, then

$E[T] = \infty, \quad E[S_T] = \frac{\sigma}{\sqrt{2}} \exp \sum_{k=1}^{\infty} \frac{1}{k}\left\{\frac{1}{2} - P[S_k > 0]\right\} < \infty$. Theorem 5 is of relevance also in the study of the queue in which the mean of the inter-arrival time equals that of the service time.

An examination of Spitzer's proofs of the results of this section shows that in every case except (3.12) and (3.4), all statements would remain valid even if the random variables X_i were continuous.

Spitzer's proof of (3.12) was based on an operator identity of Baxter (1961; section 2, part 1). An analytic proof of (3.12) using this operator was given by Baxter (1961, Ex. 3.1) for the continuous case.

For the remainder of this section, we sketch a proof of (3.12) to illustrate once again the analytic approach to fluctuation theory.

Following Baxter, we define $\mathcal{A} = \left\{\phi / \phi(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} dF(x); \quad F(x) \text{ of bounded variation}\right\}$. For any $\phi(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} dF(x)$ in $\mathcal{A}$, let $\phi^+(\theta) = \int_{0+}^{\infty} e^{i\theta x} dF(x)$ and $\phi^-(\theta) = \int_{-\infty}^{0+} e^{i\theta x} dF(x)$; and let $\mathcal{A}^+ = \left\{\phi^+ / \phi \in \mathcal{A}\right\}$ and $\mathcal{A}^- = \left\{\phi^- / \phi \in \mathcal{A}\right\}$. It is easily verified that $\mathcal{A}^+$ is closed under addition, multiplication, and multiplication by a complex constant; that $(\phi+)^+ = \phi^+$; and that

$(a\phi_1 + b\phi_2)^+ = a\phi_1^+ + b\phi_2^+$, where a and b are complex numbers. Also, ~~that~~ the operator ϕ^- possesses similar properties. Moreover, $\phi = \phi_1^+ + \phi_2^-$ implies $\phi_1^+ = \phi^+$ and $\phi_2^- = \phi^-$.

For $0 \leq s < 1$ and $0 \leq t < 1$, let

$$\phi_n(s,\theta) = \sum_{k=0}^{n} s^k E\left[e^{i\theta S_n};\ N_n^+ = k\right] \qquad n \geq 0.$$

$$\psi(s,t,\theta) = \sum_{n=0}^{\infty} t^n \phi_n(s,\theta)\ , \qquad \text{and}$$

$$\phi(\theta) = E[X_i].$$

All three functions just defined belong to $\mathcal{A}$.

Now
$$\phi_n^+(s,\theta) = \int_{0+}^{\infty} e^{i\theta S_n}\, d\left\{\sum_{k=0}^{n} s^k P\left[S_n \leq x;\ N_n^+ = k\right]\right\}$$

$$= \sum_{k=0}^{n} s^k E\left[e^{i\theta S_n};\ N_n^+ = k,\ S_n > 0\right] \qquad (3.13)$$

Similarly,
$$\phi_n^-(s,\theta) = \sum_{k=0}^{n} s^k E\left[e^{i\theta S_n};\ N_n^+ = k,\ S_n \leq 0\right] \qquad (3.14)$$

But
$$\phi(\theta)\phi_n(s,\theta) = \sum_{k=0}^{n} s^k E\left[e^{i\theta S_{n+1}};\ N_{n+1}^+ = k+1,\ S_{n+1} > 0\right]$$

$$+ \sum_{k=0}^{n} s^k E\left[e^{i\theta S_{n+1}};N_{n+1}^+ = k,\ S_{n+1} \leq 0\right]$$

(3.15)

By the uniqueness of the decomposition $\phi = \phi^+ + \phi^-$,

$$s\left\{\phi(\theta)\,\phi_n(s,\theta)\right\}^+ = \phi^+_{n+1}(s,\theta) \qquad \text{from (3.13) and (3.15)}$$

and
$$\left\{\phi(\theta)\,\phi_n(s,\theta)\right\}^- = \phi^-_{n+1}(s,\theta) \qquad \text{from (3.14) and (3.15)}.$$

Hence
$$\phi_{n+1}(s,\theta) = s\left\{\phi(\theta)\,\phi_n(s,\theta)\right\}^+ + \left\{\phi(\theta)\,\phi_n(s,\theta)\right\}^- .$$

Multiplying the last identity by t^{n+1} and summing over $n \geq 0$,

$$1 + st\,(\phi\psi)^+ + t(\phi\psi)^- = \psi, \qquad \text{whence}$$

$$\left\{\psi(1-st\phi)\right\}^+ = \left\{\psi(1-t\phi)-1\right\}^- = 0, \qquad \text{and so}$$

$$\left\{\psi(s,t,\theta)\,(1-st\phi(\theta))\right\} = \int_{-\infty}^{0+} e^{i\theta x}\,dF(x;\,s,t),$$

$$\left\{\psi(s,t,\theta)\,(1-t\phi(\theta))\right\} = \int_{0+}^{\infty} e^{i\theta x}\,dG\,(x;s,t), \qquad \text{for}$$

uniquely determined functions F and G.

Let
$$g(z;\,s,t) = \int_{-\infty}^{0+} z^x\,dF(x;s,t) \qquad |z| \geq 1$$

and
$$h(z;\,s,t) = \int_{0+}^{\infty} z^x\,dG(x;s,t) \qquad |z| \leq 1.$$

Then, for $|z| = 1$, $\quad g(z;\ s,t) = \psi(s,t,\theta)\left\{1-st\phi(\theta)\right\}$ and

$$h(z;\ s,t) = \psi(s,t,\theta)\left\{1-t\phi(\theta)\right\} .$$

It is easy to see that $g(z;\ s,t)$ is analytic in $|z| > 1$ and continuous and bounded on $|z| \geq 1$, while $h(z;\ s,t)$ is analytic in $|z| < 1$ and continuous in $|z| \leq 1$.

The rest of Spitzer's proof (page 222) may now be copied verbatim to obtain the desired identity (3.12) for the continuous case. As in the proof of (3.1) and (3.2), the crux of the remaining part of this proof is the use of the theorems of Morera and Liouville to establish the same constant value for two complex functions, one analytic in $|z| < 1$ and continuous on $|z| \leq 1$, and the other analytic in $|z| > 1$ and continuous on $|z| \geq 1$.

CHAPTER IV

Applications of Fluctuation Theory to Queueing Theory

§ 1. Introduction.

The queue to be considered in this chapter and the next is the single server queue with general input and service time distributions. Customer C_o arrives at time $t = 0$ and is served immediately. Subsequent customers are served in the order of their arrival. t_i is the length of the time interval between the arrivals of customers C_i and C_{i+1}, and s_i the service time of C_i. It is assumed that $\{u_i = s_i - t_i\}$ $i = 0, 1, \ldots$ forms a renewal process i.e. $\{u_i\}$ is a sequence of independent and identically distributed random variables. The distribution of the inter-arrival and service times are denoted by $A(x)$ and $B(x)$ respectively and their kth moments, when these exist, by a_k and b_k.

Let

$$S_n = \sum_{i=o}^{n-1} s_i, \qquad T_n = \sum_{i=o}^{n-1} t_i \quad \text{and} \quad U_n = S_n - T_n.$$

The waiting time, w_n, is the time between the arrival and commencement of service of customer C_n; and v_n is the accumulated idle time of the server in $[0, T_n]$ (T_n is the time of arrival of customer C_n). By assumption, $w_o = v_o = 0$.

By a busy period is meant the (continuous) time interval between the arrival of a customer who finds an empty queue (i.e. an idle server) and the first subsequent instant when the queue is again empty. The convention will be adopted that the busy period has not ended at an instant when the departure of the last customer in a queue coincides with the arrival of the next one.

An idle period is the time interval between the end of a busy period and the start of the next.

Let N (which may be infinite) be the number of customers served in a busy period, $N_\varepsilon(n)$ the number of idle periods in $[0, T_n]$, and $\gamma_k = P[N=k]$. We note that the length of a busy period is S_N, while that of an idle period is $-U_N$.

The study of the busy period is greatly simplified by the applicability of renewal theory arguments; for the instant of commencement of a busy period is a regeneration point of a regenerative process, the subsequent state of the queue being independent of what has occurred prior to the commencement. This is because the second busy period begins with the arrival of customer C_N, where N, the number of customers served in the first busy period, is $\min_{n\geq 1} \{n/-U_n > 0\}$. From Chapter III, N is a ladder index; hence, the instant of commencement of a busy period is a regeneration point.

If ε is the event that a customer does not have to wait for service viz. the event that a busy period begins, then ε is clearly a recurrent event, and for $n \geq 1$,

$$P(w_n = 0) = \gamma_n + \sum_{r=1}^{n} P(w_r = 0)\, \gamma_{n-r}, \tag{1.1}$$

where $\gamma_o = 0$ and $P(w_o = 0) = 1$ (Heathcote, 1964).

By considering the random walk $-U_n$, N is, in the terminology of Chapter III, T. This leads one to expect that the results of § 3, Chapter III, can be applied to obtain fundamental properties of the general, single server queue.

Before we set about using the results of the last chapter, we investigate the relationship between w_n and the variable $M_n = \max_{o \leq j \leq n} U_j$. That $w_{n+1} = (w_n + u_n)^+$, $n \geq 0$, is not hard to see (Lindley, 1952). Using this relationship, Lindley showed by induction that

$$P[w_n \leq x] = P[u_1 + \dots + u_n \leq x,\ u_2 + \dots + u_n \leq x, \dots, u_n \leq x]$$

and hence, that w_n is stochastically equivalent to M_n. $P[w_n > x]$ is therefore non-decreasing in n, whence w_n tends in distribution to a limit w which may be finite or infinite. It was also shown by Lindley (1952) that $w < \infty$ with probability one if and only if $E(u) = b_1 - a_1 < 0$. This last result is simply Theorem 1(a) of § 3, Chapter III.

From (3.3) of Chapter III,

$$\sum_{n=0}^{\infty} x^n E\left[e^{-sw_n}\right] = \exp\left\{\sum_{1}^{\infty} \frac{x^n}{n} E\left[e^{-sU_n^+}\right]\right\} \tag{1.2}$$

Differentiating (1.2) with respect to s at s = 0,

$$\sum_{n=0}^{\infty} x^n E(w_n) = \sum_{n=0}^{\infty} x^n \sum_{n=1}^{\infty} \frac{x^n}{n} E[U_n^+], \quad \text{and so}$$

$$E[w_n] = \sum_{j=1}^{n} j^{-1} E[U_j^+] \tag{1.3a}$$

Similarly,

$$\mathrm{Var}\,[w_n] = \sum_{j=1}^{n} j^{-1} E[U_j^+]^2 - \sum\sum_{\substack{r,s\leq n \\ r+s>n}} (rs)^{-1} E[U_r^+]\, E[U_s^+] \tag{1.3b}$$

To get an identity for the idle time v_n, we note that $v_n = \min\{U_o, U_1, \ldots, U_n\}$ and use (3.3) of Chapter III, with $\eta = 1$. Then,

$$\sum_{n=0}^{\infty} t^n E\left[e^{-sv_n}\right] = \exp \sum_{n=1}^{\infty} \frac{t^n}{n} E\left[e^{-sU_n^-}\right] \tag{1.4}$$

$$E\,[v_n] = \sum_{j=1}^{n} j^{-1} E\,[U_j^-] \tag{1.5a}$$

$$\text{and } \operatorname{Var}[v_n] = \sum_{j=1}^{n} j^{-1} E[U_j^-]^2 - \sum\sum_{\substack{r,s\leq n \\ r+s>n}} (rs)^{-1} E[U_r^-]\, E[U_s^-] \quad (1.5b)$$

Heathcote (1964) obtained, by direct probabilistic arguments, the equation

$$P[v_n \leq y] = q_n + \sum_{j=0}^{n} \int_0^y P\left[v_n \leq y - u\right] d\chi_{n-j}(u) \quad (1.6)$$

where $q_n = 1 - \sum_{j=1}^{n} \gamma_j$ and $\chi_n(y) = P\left[\min_{1\leq j\leq n-1} U_j \geq 0,\ -y \leq U_n < 0\right]$,

and from it, derived (1.4) using arguments similar to those of Finch (1961).

The duality between v_n and w_n merits comment. This duality, as manifested in (1.2) and (1.4), and (1.3) and (1.5) is due to the fact that while $v_n = \min\{U_o, \ldots, U_n\} = \max\{-U_o, \ldots, -U_n\}$, w_n is stochastically equivalent to $M_n = \max\{U_o, \ldots, U_n\}$. As was noted earlier, w_n tends in distribution to a finite variable w if and only if $b_1 < a_1$. Hence, by considering the dual queue with inter-arrival time s_i and service time t_i, we see that v_n tends in distribution to a finite variable v if and only if $b_1 > a_1$. More explicitly, w_n and v_n are linked by

$$w_n = U_n + v_n \quad (1.7),$$

a relationship discovered by Kingman (1962b).

By Theorem 2 of § 3, Chapter III, the number of customers in a busy period, N, is finite if and only if $b_1 \leq a_1$. If $b_1 < a_1$, then by Theorem 3,

$$E[N] = \exp \sum_{k=1}^{\infty} k^{-1} (1 - c_k) < \infty \qquad (1.8)$$

and

$$E[\text{idle period}] = E[-U_N] = (a_1 - b_1) E[N] \qquad (1.9)$$

where $c_k = P[U_k < 0]$.

The class of GI/G/1 queues may be divided into three sub-classes, according as whether $b_1 < a_1$ (ergodic), $b_1 > a_1$ (divergent) or $b_1 = a_1$.

The queue for which $b_1 = a_1$ is the least tractable of the three and little is known about it. Approximations have been found by Kingman (1964). Considering a heavy traffic situation in which $b_1 < a_1$, but $\alpha = \dfrac{a_1 - b_1}{\sigma^2}$ is very nearly zero (σ^2 is the variance of u_o and assumed finite), Kingman applied the Central Limit Theorem to show that as $\alpha \to 0$, the distribution of w is asymptotically negative exponential, with mean $\frac{1}{2} \cdot \dfrac{\sigma^2}{a_1 - b_1}$. Additional results for this queue are obtainable from Theorem 2, Theorem 4 and Theorem 5 of § 3, Chapter III. Theorem 2 and Theorem 4 assert that the busy period is finite with probability one, but the mean number of customers served in a busy period is infinite. If, moreover, $\sigma^2 < \infty$, then the mean length of an idle period, by Theorem 5, is

$$E[-U_N] = \frac{\sigma}{\sqrt{2}} \exp \sum_1^\infty \frac{1}{k} (\tfrac{1}{2} - c_k) < \infty .$$

This case $b_1 = a_1$ will not be considered in our subsequent discussion which will be restricted to the ergodic and divergent queues only.

Before proceeding to find expressions for $E(w)$ and $E(v)$ for the ergodic and divergent queues respectively, in the next two sections, we shall obtain some basic results.

From (3.1) of Chapter III, the joint distribution of an idle period and the number of customers served in a busy period is given by

$$1 - E\left[x^N e^{-sU_N}\right] = \exp\left\{- \sum_{k=1}^\infty \frac{x^k}{k} E\left[e^{-sU_k}; U_k < 0\right]\right\} . \tag{1.10}$$

The following identity for the joint distribution of N and the busy period S_N was obtained independently by Kingman (1962a) and Finch (1962).

$$1 - E\left[x^N e^{-sS_N}\right] = \exp\left\{- \sum_{k=1}^\infty \frac{x^k}{k} E\left[e^{-sS_k}; U_k < 0\right]\right\} . \tag{1.11}$$

We conclude this section by observing that if $E[-U_N] < \infty$ and $E[N_\varepsilon(n)] < \infty$, then

$$E[v_n] = E[-U_N]\, E[N_\varepsilon(n)] \quad . \tag{1.12}$$

(1.12) is little more than a direct consequence of Wald's theorem (Johnson, 1959); for if Y_i is the length of the ith idle period, then $v_n = Y_1 + \dots + Y_{N_\varepsilon(n)}$, and the event $\{N_\varepsilon(n) = k\}$ is independent of the random variables $Y_{k+1}, Y_{k+2}, \dots$.

* * * * * * *

§ 2. The ergodic queue ($b_1 < a_1$).

Unless otherwise specified, it will be assumed throughout this section that $b_1 < a_1 < \infty$. Under this condition, the busy period is finite, and by (1.8) and an application of Wald's Theorem, we have

$$E[S_N] = b_1 \exp \sum_{k=1}^{\infty} \frac{1-c_k}{k} \tag{2.1}$$

and

$$E[-U_N] = (a_1 - b_1) \exp \sum_{k=1}^{\infty} \frac{1-c_k}{k} \tag{2.2}$$

for the busy and idle periods.

Let

$$G(x) = \sum_{n=1}^{\infty} x^n \gamma_n \quad \text{and} \quad \mu_k = \sum_{n=1}^{\infty} n^k \gamma_n,$$

where γ_n, we recall, is the probability that a busy period consists of n serves. From (1.11),

$$G(x) = 1 - \exp\left\{- \sum_{k=1}^{\infty} \frac{c_k}{k} x^k\right\} \tag{2.3a}$$

(2.3a) may alternatively be written

$$G(x) = 1 - (1-x) \exp \sum_{k=1}^{\infty} \frac{1-c_k}{k} x^k \tag{2.3b}$$

Since $b_1 < a_1$, the waiting time w_n converges in distribution to a random variable w which is finite with probability one.

The rate at which this convergence occurs will be discussed in the next chapter. Here, we are primarily concerned with expressing $E(w)$ in terms of the moments μ_k.

A lemma that will be appealed to is the following .

<u>Lemma 1</u>

Let k be a non-negative integer. Then $\sum_{n=1}^{\infty} n^k(1 - c_n)$ converges if $b_{k+2} < \infty$.

The proof of the lemma uses a theorem of Heyde's (1964, Th.1) which, in our notation, states that if $Eu_o > 0$ $(b_1 > a_1)$, then $\sum_{n=1}^{\infty} n^k P[U_n \leq x]$ converges uniformly for all x in $(-\infty,\infty)$ if and only if $E(u_o^-)^{k+2} < \infty$. By considering the reversed walk $u_n^R = -u_n$, it is evident that if $E(u_o) < 0$ $(b_1 < a_1)$, then $\sum_{n=1}^{\infty} n^k P[U_n > x]$ converges uniformly for all x in $(-\infty,\infty)$ if and only if $E(u_o^+)^{k+2} < \infty$. Since $u_o^+ = \max(0, s_o - t_o)$, $u_o^+ \leq s_o$, and so $E[u_o^+]^{k+2} \leq b_{k+2} < \infty$.

Lemma 1 may be used to determine a sufficient condition for μ_k to be finite.

Lemma 2

If k is a positive integer, then $b_{k+1} < \infty$ implies $\mu_k < \infty$.

Proof: Define functions L, K and H by

$$L(x) = \sum_{n=1}^{\infty} \frac{1-c_n}{n} x^n , \qquad K(x) = e^x \qquad \text{and}$$

$$H(x) = K(L(x)) .$$

Then, by (2.3b)

$$G(x) = \sum_{n=1}^{\infty} \gamma_n x^n = 1 - (1-x) H(x) .$$

Differentiating the above equation k times,

$$\sum_{n=1}^{\infty} n(n-1) \dots (n-k+1) x^{n-k} \gamma_n = - \frac{d^k}{dx^k} \left\{ (1-x) H(x) \right\}$$

$$= k H^{(k-1)} (x) - (1-x) H^{(k)} (x) , \tag{2.4}$$

where $H^{(k)} (x) = \frac{d^k}{dx^k} H(x)$, $H^{(o)} (x) = H(x)$.

It is easily proved by induction that $H^{(k)}(x)$, $k \geq 1$, is expressible as a finite series of terms of the kind

$$\ell \exp \left(\sum_{j=1}^{\infty} \frac{1-c_j}{j} x^j \right) \prod_{q=1}^{k} \left(L^{(q)}(x) \right)^{r_q}$$

where ℓ is a real constant, and q and r_q are non-negative integers, with $r_q \leq k$. For example,

$$H^{(4)}(x) = \exp\left(\sum_{j=1}^{\infty} \frac{1-c_j}{j} x^j\right)\left\{\left(L^{(1)}(x)\right)^4 + 6\left(L^{(1)}(x)\right)^2 L^{(2)}(x) + 3\left(L^{(2)}(x)\right)^2 + 4\,L^{(1)}(x)\,L^{(3)}(x) + L^{(4)}(x)\right\} .$$

Now,

$$L^{(q)}(x) = \sum_{n=1}^{\infty} (n-1)(n-2) \ldots (n-q+1)(1-c_n)\, x^{n-q}$$

$$< \sum_{n=1}^{\infty} n^{q-1}(1-c_n)\, x^{n-q} \leq \sum_{n=1}^{\infty} n^{k-1}(1-c_n)\, x^{n-k} \quad \text{for } x \leq 1.$$

By Lemma 1,

$$L^{(q)}(1) < \infty \quad .$$

Hence,

$$\lim_{x \to 1-} H^{(k)}(x) < \infty \text{ and by (2.4), } \mu_k < \infty.$$

The converse of the lemma is not true since it is not hard to show that $b_1 < a_1 < \infty$ implies $\mu_1 < \infty$ (recall (1.8)).

Whether lemma 2 can be strengthened to read "$b_k < \infty$ implies $\mu_k < \infty$" depends, as is evident from the proof, on whether $(1-x) \sum_{n=1}^{\infty} n^{k-1} (1-c_n)\, x^{n-k}$ has a finite limit as $x \to 1-$.

It is certainly true for k = 1 since

$$\lim_{x\to 1-} (1-x) \sum_{n=1}^{\infty} (1-c_n)\, x^{n-1} = \lim_{n\to\infty} (1-c_n) = 0 \quad .$$

Corollary to lemma 2.

$$E[N] = \mu_1 = \exp \sum_{n=1}^{\infty} \frac{1-c_n}{n} < \infty \tag{2.5}$$

If $\quad b_3 < \infty,\ \mu_2 = \mu_1\left(1 + 2\sum_{n=1}^{\infty}(1-c_n)\right) < \infty \qquad (2.6)$

If $\quad b_4 < \infty,\ \mu_3 = \mu_1\left(1 + 3\sum_{n=1}^{\infty}(1-c_n) + 3\left(\sum_{n=1}^{\infty}(1-c_n)\right)^2 + \right.$

$$\left. + 3\sum_{n=1}^{\infty} n(1-c_n)\right) < \infty \tag{2.7}$$

Proof: (2.5) is the same as (1.8). (2.6) and (2.7) are obtained by differentiation at x = 1 of (2.3b).

The following theorem, which we quote without proof, is due to Heathcote (1964, Th.6).

THEOREM A.

If $b_1 > a_1$ and $a_{r+1} < \infty$, $r \geq 1$, then $\sum_{j=1}^{\infty} j^{-1} E(U_j^-)^r < \infty$.

Again, by considering the dual queue in which the service time s_n^D is t_n, and the inter-arrival time t_n^D is s_n, i.e. $U_j^D = -U_j$, we see that theorem A is clearly equivalent to

Lemma 3

In the ergodic queue ($b_1 < a_1$), if $b_{r+1} < \infty$, $r \geq 1$, then $\sum_{j=1}^{\infty} j^{-1} E[U_j^+]^r < \infty$.

This lemma leads to

THEOREM 1.

$$\text{If } b_2 < \infty, \text{ then } E[w] = \sum_{j=1}^{\infty} j^{-1} E[U_j^+] < \infty \tag{2.8}$$

$$\text{If } b_3 < \infty, \text{ then } \operatorname{Var} w = \sum_{j=1}^{\infty} j^{-1} E[U_j^+]^2 < \infty \tag{2.9}$$

Proof From (1.3a), $E[w_n] = \sum_{j=1}^{n} j^{-1} E[U_j^+]$. Since w_n is stochastically increasing, by the monotone convergence theorem and lemma 3, (2.8) follows. Similarly, (2.9) follows from (1.3b).

In Kingman's (1962b) paper, it was asserted (Th. 1) that $b_1 < a_1 < \infty$ was sufficient to ensure the finiteness of $E[w]$. This is false. The fault in Kingman's proof was discovered by Heathcote, and stemmed from a misuse of Wald's theorem in sequential analysis.

THEOREM 2.

If $b_2 < \infty$, then

$$\lim_{n\to\infty} \left\{ E[v_n] - n(a_1 - b_1) \right\} = E[w] < \infty \quad .$$

If $\mu_2 < \infty$, then

$$E[w] = (a_1 - b_1)\left\{\frac{\mu_2}{2\mu_1} + \tfrac{1}{2} - \mu_1\right\}.$$

<u>Proof</u>. From (1.7), $E[w_n] = n(b_1 - a_1) + E[v_n]$ (2.10)

and the first half of the theorem follows. (2.10) can also be deduced from (1.3a) and (1.5a), since $E[w_n] = \sum_{j=1}^{n} j^{-1} E[U_j^+]$ $= \sum_{j=1}^{n} j^{-1} E[U_j] + \sum_{j=1}^{n} j^{-1} E[U_j^-] = n(b_1 - a_1) + E[v_n]$.

The proof of the second half of the theorem relies heavily on renewal theory arguments. Let ε be the event that an idle period has just ended. ε is said to have occurred at the nth "trial" if customer C_n begins a busy period, i.e. $w_n = 0$. As is evident from the discussion on the busy period in § 1, ε is a recurrent event, and its mean recurrence time is $\mu_1 = E[N]$. The number of occurrences of ε in the first n "trials" is just $N_\varepsilon(n)$, the number of idle periods in $[o, T_n]$. We can therefore apply the renewal theorem of Feller (1949, Th. 9) to get

$$E\left[N_\varepsilon(n)\right] - \frac{n}{\mu_1} = \frac{\mu_2}{2\mu_1^{\,2}} + \frac{1}{2\mu_1} - 1 + o(1) \qquad (2.11)$$

Since the queue is ergodic, the expectation of the idle period is given by $E[-U_N] = (a_1 - b_1)\,\mu_1$ (2.2). Therefore, by (1.12) and (2.11)

$$E[v_n] - n(a_1 - b_1) = (a_1 - b_1)\left\{ \frac{\mu_2}{2\mu_1} + \frac{1}{2} - \mu_1 \right\} + o(1) \quad .$$

Using the first part of the theorem, we complete the proof.

In the following example, Theorem 2 will be used to derive the Pollaczek-Khinchin formula for $E[w]$ for the queue M/G/1.

Example 2.1

M/G/1. From Takacs (1962, pg.62),

$$G(x) = \sum_{n=1}^{\infty} \gamma_n x^n = x\, B^*\left\{\lambda - \lambda G(x)\right\} \tag{2.12}$$

where $B(x) = P[s_o \leq x]$, the service time distribution, and $B^*(s) = \int_o^\infty e^{-sx}\, dB(x)$.

$$A(x) = P[t_o \leq x] = 1 - e^{-\lambda x} \quad .$$

Since $b_1 < a_1$, $G(1) = P[N < \infty] = 1$.

Differentiating (2.12) with respect to x at $x = 1$,

$$G'(1) = B^*(\lambda - \lambda G(1)) - B^{*\prime}(\lambda - \lambda G(x))\, \lambda G'(1), \quad \text{and so}$$

$$\mu_1 = G'(1) = \frac{1}{1-\lambda b_1} \quad .$$

Similarly, by differentiating (2.12) twice,

$$\mu_2 = \frac{\mu_1 + \lambda b_1 \mu_1 + \lambda^2 b_2 \mu_1^2}{1 - \lambda b_1} .$$

$$E[w] = \frac{(a_1 - b_1)}{2} \left\{ \frac{\mu_2}{\mu_1} + 1 - 2\mu_1 \right\} \qquad \text{by Theorem 2,}$$

$$= \frac{\lambda b_2}{2(1-\lambda b_1)}$$

$$= \frac{E[s_o^2]}{2E[-u_o]} .$$

The identities of fluctuation theory may also be used to derive an expression for the probability that the waiting time is zero ~~of no waiting~~. Let $W(t)$ be the waiting time of a fictitious customer arriving at time t and call it the virtual waiting time at t. In the notation of § 2 and § 3 of Chapter II, the probability of being in an idle period, $P[W(t) = 0]$, is just $\Pi_2(t)$. From (2.3) of Chapter II,

$$\Pi_2^*(s) = \frac{X_1^*(s) - F^*(s)}{1 - F^*(s)}$$

where $X_1(x)$ is the busy period distribution and $F(x)$ the distribution of a busy cycle, a busy cycle being defined as the sum of a busy period and the subsequent idle period.

From (1.11),

$$X_1^*(s) = 1 - \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} E\left[e^{-sS_k}; U_k < 0\right]\right\} .$$

Also,

$$F^*(s) = 1 - \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} E\left[e^{-sT_k}; U_k < 0\right]\right\}$$

(Finch, 1962, Th. 1).

Hence

$$\Pi_2^*(s) = \frac{\exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} E\left[e^{-sT_k}; U_k < 0\right]\right\} - \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} E\left[e^{-sS_k}; U_k < 0\right]\right\}}{\exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} E\left[e^{-sT_k}; U_k < 0\right]\right\}}$$

As

$$E\left[e^{-sT_k}; U_k < 0\right] = \int_0^{\infty} e^{-sx} B_n(x)\, dA_n(x) \qquad \text{and}$$

$$E\left[e^{-sS_k}; U_k < 0\right] = \int_0^{\infty} e^{-sx} \left\{1 - A_n(x)\right\} dB_n(x) ,$$

$$\Pi_2^*(s) = 1 - \exp \sum_{k=1}^{\infty} \frac{1}{k}\left[\int_0^{\infty} e^{-sx} B_n(x)\, dA_n(x) - \int_0^{\infty} e^{-sx} \left\{1 - A_n(x)\right\} dB_n(x)\right]$$

* * * * * * * * *

§ 3. The Divergent Queue ($b_1 > a_1$).

This queue has been comprehensively studied by Heathcote (1964). In this very brief section, we shall merely state the obvious divergent queue equivalent of Theorem 2 of § 2 and prove an interesting result concerning $N_\varepsilon(n)$, the number of idle periods in $[0, T_n]$.

Since $P\left[w_n \leq x\right] = P\left[\max\ (U_o, \ldots, U_n) \leq x\right]$ and $P\left[v_n \leq x\right] = P\left[\min\ (U_o, \ldots, U_n) \leq x\right]$, it is easy to see that v_n is distributed as w_n in the dual queue, and vice versa. Hence Theorem 2 of the previous section becomes

THEOREM 1. Let μ'_k be the kth moment of the distribution $P\left[N^D = n\right]$, where N^D is the number of serves in a busy period of the dual queue. $\Bigg($ From (2.3a), $\sum_{n=1}^{\infty} x^n P[N^D = n] =$ ~~1 — exp~~

$$1-\exp\left\{-\sum_{n=1}^{\infty} \frac{1-c_n}{n} x^n\right\}\Bigg).$$

Then,

$$\lim_{n\to\infty} \left\{E[w_n] - n(b_1 - a_1)\right\} = E[v] < \infty \quad \text{if} \quad a_2 < \infty$$

and

$$E[v] = (b_1 - a_1)\left\{\frac{\mu_2'}{2\mu_1'} + \tfrac{1}{2} - \mu_1'\right\} \quad \text{if } \mu_2' < \infty.$$

Example 3.1

Using Example 2.1, we see that

$$E[v] = \frac{E[t_o^2]}{2E[u_o]} \qquad \text{for the queue } G/M/1.$$

THEOREM 2. As $n \to \infty$,

$$E\left[N_\varepsilon(n)\right] \to \frac{G(1)}{1-G(1)} = \mu_1' - 1 < \infty \text{ , where}$$

μ_1' is the mean number of ~~serves~~ services in a busy period of the dual queue.

Proof. First, we show that since $b_1 > a_1$, $\sum_{n=1}^{\infty} \frac{c_n}{n} < \infty$,

whence $G(1) = 1 - \exp\left\{-\sum_{n=1}^{\infty} \frac{c_n}{n}\right\} < 1$. This is easily done by

considering the dual queue and using the result that $b_1 < a_1$ implies $\sum_{n=1}^{\infty} \frac{1-c_n}{n} < \infty$. In fact $\frac{1}{1-G(1)} = \exp \sum_{n=1}^{\infty} \frac{c_n}{n} = \mu_1'$.

$$E\left[N_\varepsilon(n)\right] = \sum_{j=1}^{\infty} P\left[N_\varepsilon(n) \geq j\right]$$

$$= \sum_{j=1}^{\infty} P\left[\text{number of services in } j \text{ busy periods} \leq n+1\right]$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{n+1} \gamma_k^{(j)} \tag{3.1}$$

where $\gamma_k^{(j)}$ is the probability that exactly k customers are served in j busy periods.

$$\text{Let} \quad G^{(j)}(x) = \sum_{k=1}^{\infty} \gamma_k^{(j)} x^k .$$

Then, since the start of a busy period is a regeneration point,

$$\gamma_k^{(j)} = \sum_{r=1}^{k} \gamma_r^{(j-1)} \gamma_{k-r} \qquad (\gamma_o = 0) \quad ,$$

and so $G^{(j)}(x) = \{G(x)\}^j$.

Hence from (3.1),

$$\lim_{n\to\infty} E\left[N_\varepsilon(n)\right] = \sum_{j=1}^{\infty} G^{(j)}(1)$$

$$= \frac{G(1)}{1-G(1)} < \infty \quad .$$

Theorem 2 may also be obtained by considering the probability of no waiting, $P[w_n = 0]$. From (1.1),

$$\sum_{n=1}^{\infty} x^n P[w_n = 0] = \frac{G(x)}{1-G(x)} \quad .$$

Obviously, $E\left[N_\varepsilon(n)\right] = \sum_{k=1}^{n} P[w_k = 0]$.

CHAPTER V

Rates of Convergence of Waiting and Idle Times

§ 1. Preliminary Results.

In the ergodic queue, it is known that w_n tends in distribution to a random variable w which is finite with probability one. The rate at which this convergence takes place forms the chief object of study in this chapter.

Theorem 1, from which the major results of this chapter (Theorems 2, 3, and 4) follow, may be regarded as a straightforward corollary of Theorem 2 of Heyde (1964). However, Heyde's theorem is too general, and the methods used in deriving it unnecessarily involved, for our purposes. For in the queueing case, u_n is expressible as the difference of two positive random variables s_n and t_n, where $\{s_n\}_{n=0}^{\infty}$ and $\{t_n\}_{n=0}^{\infty}$ are assumed to be two mutually independent renewal sequences. This special property of u_n makes it possible to prove Theorem 1 by relatively simple arguments.

Heathcote has shown that in the divergent queue, the idle time v_n tends in distribution to the finite variable v exponentially fast under certain conditions. By considerations of duality, this same rate of convergence must obtain for the waiting time w_n in the ergodic queue. We opt to omit duality arguments here because in the course of establishing the exponential rate of convergence without resorting to such arguments, methods are employed and supplementary results obtained which are of interest in their own right. The methods used here resemble those of Heathcote's (1964).

It will be shown that under similar conditions, the rates of convergence of $P[w_n = 0]$, $P[v_n \leq x]$ and $E[w_n]^k$ (k any positive integer) to $P[w = 0]$, 0 and $E[w]^k$ respectively are also exponential.

Before proceeding to prove the main theorems in § 2, we shall first derive some preliminary results.

Throughout this chapter, $\int_{-\infty}^{\infty} e^{-sx} \, dF(x)$ will be denoted by $F^*(s)$ and the convolution

$$\int_0^t F(t-u) \, dG(u)$$

by ~~F*G(t)~~ $F \ast G(t)$. Note that

$$1-c_n = P[U_n \geq 0] = \int_0^\infty \left\{1-B_n(y)\right\} dA_n(y) = \int_0^\infty A_n(y) \, dB_n(y).$$

As in the proof of Theorem 2 of § 2, Chapter IV, if $\mathcal{E}$ is the event that an idle period has just ended, and $\mathcal{E}$ is said to occur at the nth trial if customer C_n does not have to wait for service, then $\mathcal{E}$ is a recurrent event and the probability that $\mathcal{E}$ occurs at the nth trial is $P[w_n = 0]$. By the elementary renewal theorem of Feller (1949),

$$\lim_{n\to\infty} P[w_n = 0] = P[w = 0] = \frac{1}{\mu_1} \tag{1.1}$$

where μ_1, the mean recurrence time of $\mathcal{E}$, is given by

$$\mu_1 = E[N] = \exp \sum_{k=1}^{\infty} \frac{1-c_k}{k} .$$

By the corollary to Lemma 2 of § 2, Chapter IV, $\mu_1 < \infty$ if $b_1 < a_1$. $\frac{1}{\mu_1}$ is interpreted as zero if $\mu_1 = \infty$ i.e. if $b_1 = a_1$, or if $P[N < \infty] < 1$ i.e. if $b_1 > a_1$.

A similar renewal type argument yields a representation for $P[w_n \leq y]$ when $y \geq 0$. If customer C_n joins a non-empty queue, his waiting time w_n is independent of the process prior to the initiation of the current busy period. Conditional on this busy period commencing on the arrival of customer C_{n-k}, it follows from the recurrence relation $w_{n+1} = (w_n + u_n)^+$ that $w_n = \sum_{i=o}^{k-1} u_{n-k+i}$, so that w_n, conditionally, has the same distribution as U_k^+. Hence,

$$P[w_n \leq y] = \sum_{k=1}^{n} P[U_k^+ \leq y] \; P[w_{n-k} = 0, \; w_\nu > 0 \text{ for all } n\text{-}k<\nu<n].$$

The second probability on the right is simply the probability that on the arrival of customer C_n, the last occurrence of the recurrent event ε is exactly k trials away. In terms of the tail probability $q_{k-1} = \sum_{j=k}^{\infty} \gamma_j$,

$$P[w_{n-\kappa} = 0, \; w_\nu > 0 \text{ for all } n\text{-}k < \nu < n] = q_{k-1} \, P[w_{n-k} = 0].$$

These results can be summed up by

Lemma 1

For $y \geq 0$,

$$P[w_n \leq y] = \sum_{k=1}^{n} P[U_k^+ \leq y] \, q_{k-1} \, P[w_{n-k} = 0] \, . \qquad (1.2)$$

If $b_1 < a_1$ so that the queue is ergodic,

$$P[w \leq y] = \lim_{n\to\infty} P[w_n \leq y] = \frac{1}{\mu_1} \sum_{k=1}^{\infty} P[U_k^+ \leq y] \, q_{k-1} \, . \qquad (1.3)$$

In deducing (1.3) we have used the easily verified result that if $\sum_{k=1}^{\infty} a_k < \infty$ and $b_k \to b$, then $\sum_{k=1}^{n} a_k b_{n-k} \to b \sum_{k=1}^{\infty} a_k$.

Note that w is indeed an honest random variable, as $P[w < \infty] = \frac{1}{\mu_1} \sum_{k=1}^{\infty} q_{k-1} = 1$.

There remains to be proved one other Lemma, which is also of some independent interest.

Lemma 2

Suppose $F(t)$ is a distribution function on $[0,\infty)$ with finite mean m_1. Let $K(z,t) = \sum_{n=1}^{\infty} z^n F_n(t)$, where $F_n(t)$ is the nth iterated convolution of $F(t)$. If $F(0) < 1$, then there exists an $x_o > 1$ such that, as $t \to \infty$,

$$K(x_o,t) = O(e^{\epsilon t}),$$

where ϵ is the unique positive root in s of the equation $x_o F^*(s) = 1$.

Proof. Note briefly that by Theorem 1 of Belyaev and Maksimov (1963), $K(z,t)$ is analytic inside a circle of radius $[F(0)]^{-1}$ for all finite t. For a fixed x_o in the interval $(1, [F(0)]^{-1})$, the convexity of $F^*(s)$ and the condition of $F^*(0) = 1$ establishes that the equation $x_o F^*(s) = 1$ has a unique root in s, say $s = \epsilon$.

Now,

$$e^{-\epsilon t}K(x_o,t) = x_o e^{-\epsilon t}F(t) + x_o e^{-\epsilon t}\left\{F(t) * \sum_{n=2}^{\infty} x_o^{n-1}F_{n-1}(t)\right\}$$

$$= x_o e^{-\epsilon t} F(t) + \int_o^t e^{-\epsilon(t-y)}K(x_o,t-y)\, x_o e^{-\epsilon y}dF(y)$$

where $*$ denotes the convolution operator, i.e. $A(t) * B(t) = \int_o^t A(t-u)\, dB(u)$.

Let $G(t) = \int_o^t x_o e^{-\epsilon y} dF(y)$. $G(t)$ is positive, non-decreasing, has total variation unity, and so is an honest distribution function. Further, $G(t)$ has finite mean since

$$\int_o^t t\, dG(t) = \int_o^t t\, x_o e^{-\epsilon t} dF(t) \leq x_o m_1 < \infty .$$

If we define $H(x_o,t) = \int_o^t e^{-\epsilon y} K(x_o,y)\, dy$, then it is easy to see that $H(x_o,t)$ satisfies the renewal type equation

$$H(x_o,t) = \int_o^t x_o e^{-\epsilon y} F(y)\, dy + \int_o^t H(x_o,t-y)\, dG(y).$$

By the elementary renewal ~~th eorem~~ theorem (Smith, 1954),

$$\lim_{t\to\infty} \frac{1}{t} H(x_o,t) = \frac{\int_o^\infty x_o e^{-\epsilon t} F(t)\, dt}{\int_o^\infty t\, dG(t)} < \infty .$$

Hence $e^{-\epsilon t} K(x_o,t) = O(1)$ and the proof is complete.

* * * * * * *

§ 2. Convergence Rates.

It is assumed throughout this section that $b_1 < a_1 < \infty$.

THEOREM 1. $G(x) = \sum_{n=1}^{\infty} x^n \gamma_n$ has a radius of convergence greater than unity if and only if there exists an $\epsilon > 0$ such that $B^*(-\epsilon) < \infty$.

Proof. If $B^*(-\epsilon)$ converges for an $\epsilon > 0$ then

$$\phi(\theta) = \int_{-\infty}^{\infty} e^{\theta y} \, dP[u \leq y] = A^*(\theta) \, B^*(-\theta)$$

is defined and continuous in the interval $[0,\epsilon]$. Since $\phi'(0) = b_1 - a_1 < 0$, it follows that there exists a δ in $[0,\epsilon]$ such that $\phi(\delta) < 1$. Further,

$$\begin{aligned} [B^*(-\delta)]^n &= \int_0^{\infty} e^{\delta y} \, dB_n(y) \\ &\geq \int_y^{\infty} e^{\delta u} \, dB_n(u) \\ &\geq e^{\delta y} \, [1 - B_n(y)] \text{ for all } y \text{ in } [0,\infty) \text{ and} \end{aligned}$$

a δ in $[0,\epsilon]$. Thus

$$\begin{aligned} \sum_1^{\infty} n^{-1} x^n (1-c_n) &= \sum_1^{\infty} n^{-1} x^n \int_0^{\infty} [1-B_n(y)] \, dA_n(y) \\ &\leq \sum_1^{\infty} n^{-1} x^n \int_0^{\infty} e^{-\delta y} \, [B^*(-\delta)]^n \, dA_n(y) \\ &= - \log [1 - x \, B^*(-\delta) \, A^*(\delta)]. \end{aligned}$$

The series on the left has a radius of convergence greater than unity because $B^*(-\delta)\ A^*(\delta) < 1$ for some δ in $[0,\epsilon]$. From (2.3b) of Chapter IV it follows that the series $G(x)$ converges for an $x > 1$.

Conversely, if $G(x)$ and hence $\sum_1^\infty n^{-1}\ x^n(1-c_n)$ has a radius of convergence $r > 1$, then the series of derivatives $\sum_1^\infty x^n(1-c_n)$ also converges for $|x| < r$. Thus there exists an $x_o > 1$ such that

$$\infty > \sum_1^\infty x_o^n\ (1-c_n) = \sum_1^\infty x_o^n \int_o^\infty [1-B_n(y)]\ dA_n(y)$$

$$\geq \sum_1^\infty x_o^n \int_o^\infty [1-B(y)]\ dA_n(y)$$

$$= \int_o^\infty \sum_1^\infty x_o^n\ A_n(y)\ dB(y).$$

By the theorem of Belyaev and Maksimov (1963), there exists a $y_o > 1$ such that $\sum_1^\infty y_o^n\ A_n(t) < \infty$ for all finite t. Taking $z_o = \min\ (x_o,\ y_o)$ we have

$$\infty > \int_o^\infty \sum_1^\infty\ z_o^n\ A_n(y)\ dB(y),$$

and the result of Lemma 2 implies that the integrals

$$\int_o^\infty \sum_1^\infty z_o^n\ A_n(y)\ dB(y) \text{ and } \int_o^\infty e^{\epsilon y}\ dB(y)$$

converge or diverge together for an $\epsilon > 0$. In fact, ϵ is the root in s of the equation $z_oA^*(s) = 1$.

The simplicity of the above proof is due largely to the expressibility of u_n as the difference of two positive, mutually independent random variables.

Theorem 1 in conjunction with Lemma 1 is useful in studying the rate of convergence of $P[w_n \leq y]$ to its limit. But first it is convenient to prove

<u>THEOREM 2</u>. There exists an $\alpha < 1$ such that

$$\lim_{n\to\infty} \frac{P[w_n = 0] - P[w = 0]}{\alpha^n} = 0$$

if and only if there exists an $\epsilon > 0$ such that $B^*(-\epsilon) < \infty$.

<u>Proof</u>. From (1.1) of Chapter IV and (1.1), for $n = 1, 2, \ldots,$

$$P[w_n = 0] - P[w = 0] = \sum_{r=o}^{n} \left\{P[w_r=0] - P[w=0]\right\} \gamma_{n-r} +$$

$$+ P[w=0] \sum_{r=o}^{n} \gamma_{n-r} - P[w=0]$$

$$= \sum_{r=o}^{n} \left\{P[w_r=0] - P[w=0]\right\} \gamma_{n-r} - \mu_1^{-1} \sum_{r=n+1}^{\infty} \gamma_r .$$

Let $Q(x) = \sum_{n=o}^{\infty} x^n \left\{P[w_n = 0] - P[w = 0]\right\}.$

Since $P[w_o = 0] = 1$ and $\gamma_o = 0$,

$$Q(x) - \left\{1-P[w=0]\right\} = \sum_{n=1}^{\infty} x^n \sum_{r=o}^{n} \left\{P[w_r=0] - P[w=0]\right\} \gamma_{n-r} - \mu_1^{-1} \sum_{n=1}^{\infty} x^n \sum_{r=n+1}^{\infty} \gamma_r,$$

and

$$Q(x) = 1 + \sum_{n=o}^{\infty} x^n \sum_{r=o}^{n} \left\{P[w_r=o] - P[w=0]\right\} \gamma_{n-r} - \mu_1^{-1} \sum_{n=o}^{\infty} x^n \sum_{r=n+1}^{\infty} \gamma_r$$

$$= 1 + Q(x)\, G(x) - \frac{G(1) - G(x)}{\mu_1(1-x)}$$

$$= \frac{1}{1-G(x)} - \frac{1}{\mu_1(1-x)} \quad .$$

From (2.3b) of Chapter IV,

$$Q(x) = \frac{\mu_1 \exp\left\{- \sum_1^{\infty} n^{-1} x^n (1-c_n)\right\} - 1}{\mu_1(1-x)} \tag{2.1}$$

and

$$Q(1) = \frac{\frac{d}{dx}\left[\exp\left\{- \sum_1^{\infty} n^{-1} x^n (1-c_n)\right\}\right]_{x=1}}{\frac{d}{dx}\left[1-x\right]_{x=1}}$$

$$= \mu_1^{-1} \sum_1^{\infty} (1-c_n) \quad .$$

Suppose now that $B^*(-\epsilon) < \infty$ for an $\epsilon > 0$. By the previous theorem this implies that the radius of convergence r of $\sum_1^{\infty} n^{-1} x^n (1-c_n)$ is greater than unity. Thus $Q(1) = \mu_1^{-1} \sum_1^{\infty} (1-c_n) < \infty$ and so by (2.1), $Q(x)$ has a radius of convergence $r > 1$. Choosing α so that $r^{-1} < \alpha < 1$ we have also the convergence of the series $Q\left(\frac{1}{\alpha}\right)$, implying that

$$\alpha^{-n}\left\{P[w_n=0] - P[w=0]\right\} \to 0 \quad \text{as} \quad n \to \infty.$$

Conversely, $P[w_n=0] - P[w=0] = o(\alpha^n)$ for some $\alpha < 1$ implies that $Q(x)$, and hence by (2.1), $\sum_1^\infty n^{-1} x^n (1-c_n)$, converges for an $x > 1$, and an appeal to Theorem 1 completes the proof. (If x is in $(1,\alpha^{-1})$, then $Q(x) = \sum_{n=0}^\infty (x\alpha)^n \dfrac{\{P[w_n=0] - P[w=0]\}}{\alpha^n} < \infty$).

<u>THEOREM 3</u>. There exists an $\alpha < 1$ such that

$$\lim_{n\to\infty} \frac{P[w_n \leq y] - P[w \leq y]}{\alpha^n} = 0$$

if and only if there exists an $\epsilon > 0$ such that $B^*(-\epsilon) < \infty$. When the asserted result holds, it is true uniformly for $y \geq 0$.

<u>Proof</u>. Let $R(x,y) = \sum_{n=1}^\infty x^n q_{n-1} P[U_n^+ \leq y]$. From Lemma 1 and (1.1)

$$P[w_n \leq y] - P[w \leq y] = \sum_{k=1}^n q_{k-1} P[U_k^+ \leq y] \left\{P[w_{n-k}=0] - P[w=0]\right\}$$

$$- P[w=0] \sum_{k=n+1}^\infty q_{k-1} P[U_k^+ \leq y] \quad ,$$

and hence

$$\sum_{n=1}^\infty x^n \left\{P[w_n \leq y] - P[w \leq y]\right\} = R(x,y)Q(x) - \frac{\{xR(1,y) - R(x,y)\}}{\mu_1(1-x)} \qquad (2.2)$$

Further,

$$R(x,y) \leq \sum_{1}^{\infty} x^n q_{n-1} = \frac{x\{1-G(x)\}}{1-x}$$

$$\rightarrow \frac{1}{\mu_1} \text{ as } x \rightarrow 1-, \quad \text{by (2.3b) of Chapter IV.}$$

If we now suppose that $B^*(-\epsilon) < \infty$ for an $\epsilon > 0$, then by Theorem 1, $R(x,y)$ has a radius of convergence greater than unity for all $y \geq 0$, and by Theorem 2, so has $Q(x)$. Thus the radius of convergence R of the series on the left side of (2.2) is greater than unity. (It is easily checked that

$$\lim_{x \to 1} \frac{\{xR(1,y) - R(x,y)\}}{\mu_1(1-x)} < \infty \text{).}$$

Selecting an α in the interval $(R^{-1}, 1)$ we have that

$\sum_{1}^{\infty} \alpha^{-n} \{P[w_n \leq y] - P[w \leq y]\}$ converges uniformly in $y \geq 0$ and

hence $\alpha^{-n} \{P[w_n \leq y] - P[w \leq y]\} \rightarrow 0$ as asserted.

The converse follows immediately since if the right side of (2.2), and hence $Q(x)$, converges for an $x > 1$ then by Theorem 2, $B^*(-\epsilon) < \infty$ for an $\epsilon > 0$.

<u>THEOREM 4.</u> Let r be a positive integer. There exists an $\alpha < 1$ such that $\dfrac{E[w]^r - E[w_n]^r}{\alpha^n} \rightarrow 0$ as $n \rightarrow \infty$ if and only if there exists an $\epsilon > 0$ such that $B^*(-\epsilon) < \infty$.

<u>Proof</u>. Suppose $B^*(-\epsilon) < \infty$ for some $\epsilon > 0$. Then all moments b_k are finite and it is easy to show (see (1.3a) and (1.3b) of Chapter IV) that $E[w_n]^k \leq E[w]^k < \infty$. Hence

$$E[w]^k - E[w_n]^k = \int_0^\infty ky^{k-1} \left\{ P[w_n \leq y] - P[w \leq y] \right\} dy < \infty,$$

so that with $k = 2r-1$, $y^{2r-2}\left\{P[w_n \leq y] - P[w \leq y]\right\} \to 0$

as $y \to \infty$. Further, by Theorem 3, there is an $\alpha < 1$ such that

$$\frac{P[w_n \leq y] - P[w \leq y]}{\alpha^{2n}} \to 0 \text{ as } n \to \infty \quad ,$$

uniformly in y. Hence there exists k such that

$$\frac{y^{r-1} \{P[w_n \leq y] - P[w \leq y]\}}{\alpha^n} \to 0$$

uniformly in y for all $y \geq k$, as $n \to \infty$. Now,

$$\lim_{n\to\infty} \frac{E[w]^r - E[w_n]^r}{\alpha^n} = \lim_{n\to\infty} \int_0^\infty \frac{ry^{r-1} \{P[w_n \leq y] - P[w \leq y]\}}{\alpha^n} dy$$

$$= r \int_0^\infty \lim_{n\to\infty} \frac{y^{r-1}\{P[w_n \leq y] - P[w \leq y]\}}{\alpha^n} dy$$

$$= 0$$

which proves the sufficiency part of the theorem. The converse follows easily from the last few equations and Theorem 3.

<u>THEOREM 5.</u> There exists an $\alpha < 1$ such that

$$\frac{P[v_n \leq y]}{\alpha^n} \quad \to \quad 0 \qquad \text{uniformly in } y$$

if and only if there exists an $\epsilon > 0$ such that $B^*(-\epsilon) < \infty$.

<u>Proof</u>. We shall first derive an expression for $P[v_n \leq x]$ similar to that for $P[w_n \leq x]$ (1.2). Then, with the aid of Theorem B of Heyde (1964), the proof proceeds along the same lines as those of the preceeding four theorems. As in Lemma 1, we use the recurrent property of the points of commencement of busy periods. The idle time of the server at the time of arrival of customer C_n is $-U_M$, where M is the index of the customer who starts the current busy period. So,

$$P[v_n \leq y] = \sum_{k=o}^{n} P[-U_M \leq y;\ M = k]$$

$$= \sum_{k=o}^{n} P[-U_k \leq y]\ P[\mathbf{v}_k = 0;\ w_\nu > 0 \text{ for all } k < \nu \leq n]$$

$$= \sum_{k=o}^{n} P[-U_k \leq y]\ P[w_k = 0]\ q_{n-k} \quad .$$

Let

$$S(x,y) = \sum_{k=o}^{\infty} x^k\ P[-U_k \leq y]\ P[w_k = 0]$$

Then,

$$\sum_{n=0}^{\infty} x^n P[v_n \leq y] = S(x,y) \sum_{n=0}^{\infty} x^n q_n$$

$$= \frac{S(x,y)\{1-G(x)\}}{1-x} \tag{2.3}$$

We now use a result of Heyde's (1964, Theorem 2) which, in our notation, asserts that if $E[-u_o] < \infty$ and $E[-u_o] > 0$ (i.e. $b_1 < a_1$), then $\sum_{n=1}^{\infty} e^{rn} P[-U_n \leq y]$ converges for all y in $(-\infty,\infty)$ and some $r > 0$ if and only if $(-u_o)^- = u_o^+$ has an analytic characteristic function.

Suppose $B^*(-\epsilon) < \infty$ for some $\epsilon > 0$. Then $B^*(s)$ is analytic. As $u_o^+ = \max(0, s_o - t_o) \leq s_o$, u_o^+ has an analytic characteristic function. Hence, by Theorem 2 of Heyde, there exists $r > 0$ such that $\sum_{n=1}^{\infty} e^{rn} P[-U_n \leq y] < \infty$ for all y in $(-\infty,\infty)$. It now follows that $S(x,y) \leq \sum_{n=0}^{\infty} x^n P[-U_n \leq y]$ converges for some $x_o > 1$. By (2.3) and Theorem 1, $\sum_{n=0}^{\infty} x^n P[v_n \leq y]$ has a radius of convergence greater than unity whence, by the usual argument, there exists an $\alpha < 1$ such that $\frac{P[v_n \leq y]}{\alpha^n} \to 0$ uniformly in y for $y \geq 0$.

Conversely, if $P[v_n \leq y] = o(\alpha^n)$ for some $\alpha < 1$, then $G(x)$ has a radius of convergence greater than unity and using Theorem 1, we complete the proof.

When the distribution $B(x)$ has only a finite number of moments these theorems fail and the rate of convergence of $P[w_n = 0]$, $P[w_n \leq y]$ etc. is of the order of some power of n, the power depending on the number of moments that are finite. This situation can be investigated by methods similar to those used above.

THEOREM 6. If r is a positive integer and $b_{r+2} = \int_0^\infty x^{r+2}\, dB(x) < \infty$, then as $n \to \infty$

$$P[w_n = 0] - P[w = 0] = o(n^{-r}) \quad .$$

Proof. In the course of proving Theorem 2, we obtained

$$Q(x) = \sum_{n=0}^{\infty} x^n \left\{ P[w_n = 0] - P[w = 0] \right\}$$

$$= 1 + Q(x)\, G(x) - \frac{\{1-G(x)\}}{\mu_1 (1-x)} \qquad (2.4)$$

Clearly, it suffices to prove that $Q^{(r)}(1) < \infty$, where $Q^{(r)}(x) = \dfrac{d^r Q(x)}{dx^r}$. It was shown in the proof of Lemma 2 of Chapter 4 that $b_{r+2} < \infty$ implies

$$\left[\frac{d^{r+1}}{dx^{r+1}} \left\{ (1-x)^{-1} \, (1-G(x)) \right\} \right]_{x=1} = k < \infty .$$

Differentiating (2.4) (r+1) times at x = 1,

$$Q^{(r+1)}(x) = \sum_{j=o}^{r+1} \binom{r+1}{j} Q^{(j)}(1)\, G^{(r+1-j)}(1) - \mu_1^{-1} k,$$

and so

$$(r+1)\, \mu_1\, Q^{(r)}(1) = \mu_1^{-1} k - \sum_{j=o}^{r-1} \binom{r+1}{j} Q^{(j)}(1)\, G^{(r+1-j)}(1).$$

The proof can now be completed by the use of Lemma 2 of Chapter 4 and mathematical induction. It is not hard to show from (2.4) that $Q^{(1)}(1) < \infty$.

A question that remains to be answered is whether Theorems 1 - 5 are best possible. The final theorem of this chapter asserts that they are. We consider first an example.

Example 2.1.

For M/M/1,

$$G(x) = \sum_1^\infty x^n \gamma_n = (1+\rho)(2\rho)^{-1} \left\{1 - \sqrt{1-4\rho x\ (1+\rho)^{-2}}\right\}$$

where $\rho = b_1\, a_1^{-1}$ is the traffic intensity (Takács 1962, pg.31). The radius of convergence r of $G(x)$ is $(1+\rho)^2\ (4\rho)^{-1} > 1$ for $\rho \neq 1$. $G(r) = (1+\rho)(2\rho)^{-1}$ and when $\rho < 1$, $G(1) = 1$. Also, $\mu_1 = G'(1) = (1-\rho)^{-1}$. In the interval $[0,r]$ $G(x)$ is strictly increasing and attains its maximum value at $x = r$.

By (2.1),

$$Q(r) = \frac{1}{1-G(r)} - \frac{1}{\mu_1(1-r)} = \frac{2\rho}{1-\rho} < \infty \text{ for all } \rho < 1.$$

The best possible value of α in Theorem 2 is therefore

$$\alpha = r^{-1} = 4\rho(1+\rho)^{-2} .$$

This simple example leads one to the following result.

THEOREM 7. If r is the radius of convergence of $G(x) = \sum_{1}^{\infty} x^n \gamma_n$, $r > 1$ and $G(r)$ is convergent, then there does not exist a number β in $[r^{-1}, 1)$ such that

$$\frac{P[w_n = 0] - P[w = 0]}{\beta^n} \to \ell > 0 .$$

Proof. Since $G(1) = 1$ and $G(x)$ is strictly increasing, $G(r) > 1$.

By (2.1) $Q(r) < \infty$. Hence

$$\frac{P[w_n = 0] - P[w = 0]}{(r^{-1})^n} \to 0, \text{ and by comparison,}$$

$$\frac{P[w_n = 0] - P[w = 0]}{\beta^n} \to 0, \text{ for all } \beta \text{ in } [r^{-1},1).$$

Under the hypotheses of the theorem, the result

$$\frac{P[w_n = 0] - P[w = 0]}{(r^{-1})^n} \rightarrow 0,$$ is therefore the best possible.

Similarly, whenever G(r) is convergent, Theorem 3 cannot be improved upon.

REFERENCES

1. Andersen, E.S. (1953) "On sums of symmetrically dependent random variables", Skandinavisk Aktuarietidskrift, **36**, 123-138

2. ---- (1954) "On the fluctuations of sums of random variables II", Math. Scand., **2**, 195-223.

3. Baxter, G. (1961) "An analytic approach to finite fluctuation problems in probability", J. Anal. Math., **9**, 31-69.

4. Belyaev, Ya. K. and Maksimov, V.M. (1963) "Analytic properties of a generating function for the number of renewals", Theory of Probability (SIAM Translation), **8**, 102-104.

5. Cox, D.R. (1962) "Renewal Theory" Methuen, London.

6. Cox, D.R. and Smith, W.L. (1961) "Queues" Methuen, London.

7. Feller, W. (1949) "Fluctuation Theory of recurrent events", Transactions Amer. Math. Soc., **67**, 98-119.

8. ---- (1959) "On combinatorial methods in fluctuation theory", Harold Cramér volume, 75-91. John Wiley, New York.

9. Finch, P.D. (1961) "On the busy period in the queueing system GI/G/1", J. Aust. Math. Soc., **2**, 217-228.

10. Heathcote, C.R. (1964) "Divergent single server queues", To appear in the Proceedings of the Symposium on Congestion Theory held at Chapel Hill.

11. Heyde, C.C. (1964) "Two probability theorems and their application to some first passage problems", J. Aust. Math. Soc., **4**, 214 - 222.

12. Johnson, N.L. (1959) "A proof of Wald's theorem on cumulative sums", Ann. Math. Statist., **30**, 1245-1247.

13. Kingman, J.F.C., (1962a) "The use of Spitzer's identity in the investigation of the busy period and other quantities in the queue GI/G/1", J. Aust. Math. Soc., **2**, 345-355.

14. ---- (1962b) "Some inequalities for the queue GI/G/1", Biometrika, **49**,315-324.

15. ---- (1964) "The heavy traffic approximation in the theory of queues", To appear in the Proceedings of the Symposium on Congestion Theory held at Chapel Hill.

16. Lindley, D.V. (1952) "The theory of queues with a single server", Proc. Camb. Phil. Soc., **48**, 277-289.

17. Port, C.S. (1963) "An elementary probability approach to fluctuation theory", J. Math. Analysis and Applications, **6**, 109-151.

18. Smith, W.L. (1954) "Asymptotic renewal theorems", Proc. Royal Soc. Edin. series A, **64**, 9-48.

19. ---- (1955) "Regenerative stochastic processes", Proc. Royal Soc., series A, **232**, 6-31.

20. Spitzer, F. (1956) "A combinatorial Lemma and its application to probability theory", Transactions Amer. Math. Soc., **82**, 323-339.

21. ---- (1964) "Principles of random walk", van Nostrand, Princeton, N.J.

22. Takács, L. (1962) "Introduction to the Theory of Queues", Oxford U.P., New York.

ACKNOWLEDGEMENTS

I am indebted to the Colombo Plan for the research fellowship that made this work possible, and to Mrs. S. McKeown for typing assistance.

My special thanks go to my supervisor, Dr. C.R.Heathcote for his many helpful suggestions, his encouragement and infinite patience.